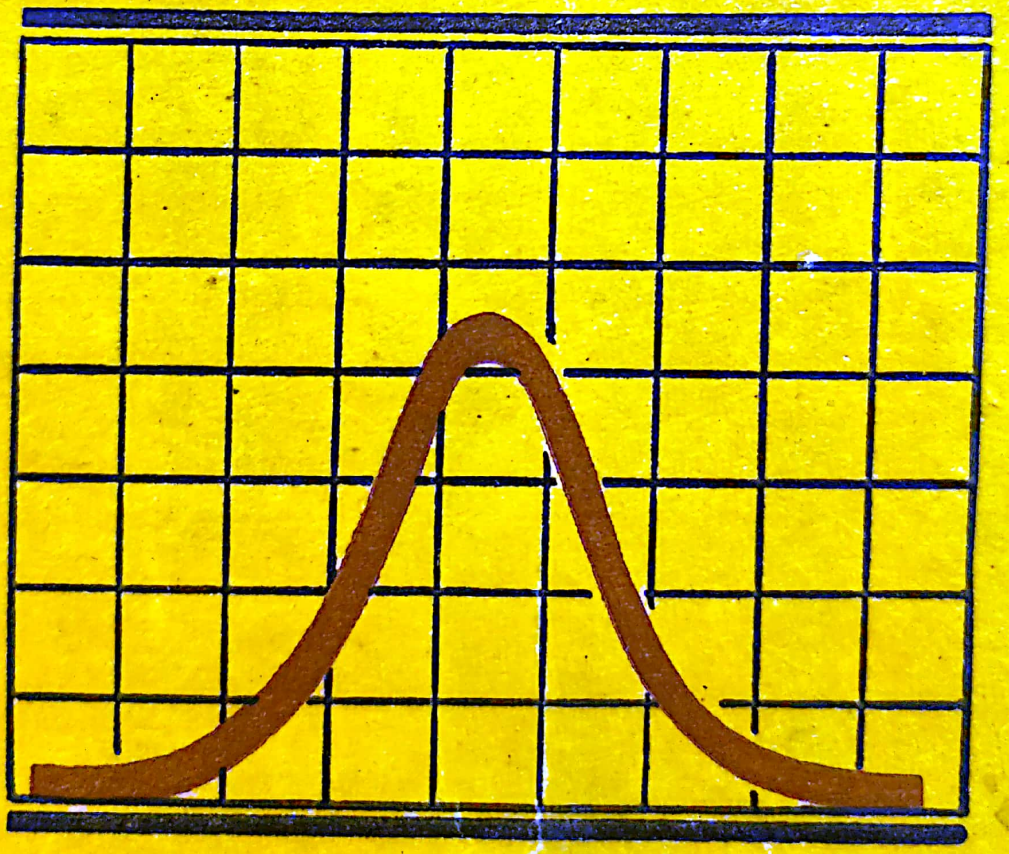
Ioan Văleanu • Maria Hincu

ELEMENTE DE STATISTICĂ GENERALĂ



Editura Litera

Dr. Ioan Văleanu
medic primar bacteriolog,
medic primar igienist,
radioizotopist

Dr. Hîncu Maria
medic principal laborator clinic
și microbiologie

ELEMENTE DE STATISTICĂ GENERALĂ

Editura Litera, București — 1990

Scanned with OKEN Scanner

ISBN 973-43-0032-6

Scanned with OKEN Scanner

Ignoranța nu este un argument
(SPINOZA)

Prefață

Cunoașterea și aplicarea unor calcule statistice este absolut necesară, deoarece orice studiu sau cercetare științifică trebuie documentată cu analiza statistică a rezultatelor.

Lucrarea încearcă a face accesibil un domeniu destul de arid, cel al formulelor matematice, putînd fi considerată un „abecedar" de statistică.

Ținînd seama că în țara noastră nu există o astfel de „culegere" de metode simple la dispoziția începătorilor în acest domeniu (și începători vor exista mereu), lucrarea va fi utilă tuturor celor ce doresc să-și verifice acuratețea cercetărilor, indiferent de specialitate.

Exemplele sînt simple, clare, de înțeles pentru orice persoană cu pregătire matematică medie, calculele putînd fi utilizate cu sau fără calculator (de buzunar), formulele fiind selectate astfel încît să fie ușor de aplicat. Cifrele, datele, exemplele au doar un rol didactic, neputînd fi luate ca surse de referință.

Lucrarea va fi utilă medicilor umani, veterinari, farmaciștilor, chimiștilor, biologilor, agronomilor etc., precum și studenților și specialiștilor din institutele de cercetare din diferite domenii.

Scanned with OKEN Scanner

Capitolul 1

Introducere. Istoric.

Deoarece limbajul matematic este precis, impersonal, universal și internațional, utilizarea lui s-a impus și în aprecierea fenomenelor biologice, căci „primul pas în progresul unei științe constă în măsurarea cantităților" (Maxwell). Acest lucru are avantajul că prin abstracții matematice se poate ajunge cu gîndirea rațională dincolo de gîndirea intuitivă.

Statistica este o știință utilizată „de toată lumea", de ex.:

— cantitatea de alimente și băuturi preambalate corespunde statistic cu cantitatea înscrisă pe ambalaj (de ex. pe unele sticle cu băuturi alcoolice scrie: 1 000±20 ml, ceea ce înseamnă că în sticlă poate fi 980 . . . 1 020 ml);

— foile de temperatură a bolnavilor sînt reprezentări grafice statistice;

— înnegrirea emulsiei fotografice se face logaritmic (Fig. 1), exponometrele fotografice fiind gradate logaritmic;

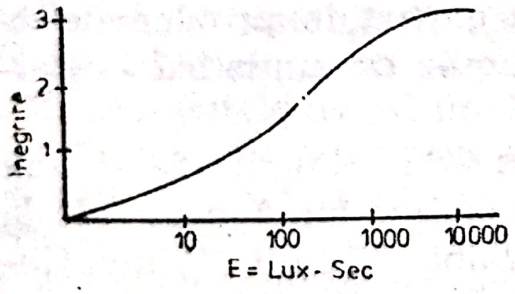


FIG. 1.

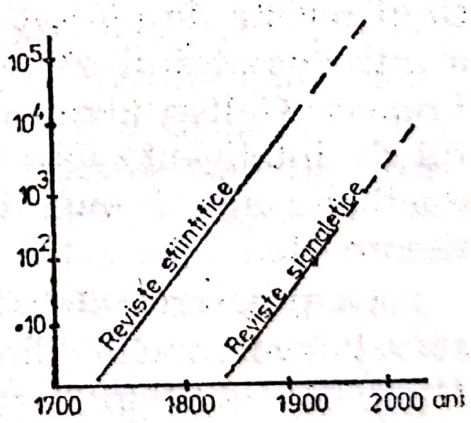


FIG. 2.

— deoarece cantitatea de informații științifice se dublează la aprox. 10 ani (după „legea" lui De Solla Price), numărul revistelor științifice și signaletice crește logaritmic (Fig. 2), din reprezentarea grafică putîndu-se prevedea statistic creșterea lor în viitor;

— adaptarea urechii la zgomot se face logaritmic;

— adaptarea ochiului la lumină se face logaritmic;
— cîștigul la loterie este o probabilitate statistică etc.

Statistica a fost supusă în decursul timpului la diverse critici și ironii („este cea mai rafinată formă a minciunii", iar Benjamin Disraeli, devenit lord Beaconsfield, fost prim-ministru al Angliei sub regina Victoria spunea că „există minciuni simple, minciuni sfruntate și statistici"). Dar, după cum a semnalat Waleix în 1840, „statistica a rezistat atît atacurilor adversarilor cît și timpului" și în orice caz, mînuită bine, ne ferește de erori.

Observarea statistică a diverselor fenomene sau boli s-a făcut din antichitate („au făcut proză fără să știe"), dar observații sistematice au făcut puțini oameni, dintre care cităm: Vitruviu, Leonardo da Vinci, Servet, Lavoisier etc., primele statistici adevărate fiind făcute în sec. XVII de John Grownt, care a făcut primele recensămînturi de bolnavi în timpul unei epidemii de ciumă din Londra, arătînd interesul ce-l prezintă construirea unor tabele de natalitate și mortalitate. De asemenea, Achenwald, matematician german, a creat termenul de „biostatistică" în 1772; Quetelet, matematician și naturalist belgian, a publicat cartea „Despre om și facultățile sale, sau studiu de fizică socială", lucrare fundamentală de biostatistică; Karl Pearson, matematician și statistician englez, a condus mult timp revista de biostatistică „Biometrika" și în anul 1900 a introdus „testul χ^2", de largă aplicabilitate în statistică; Francisc Galton (despre care se spune că a avut coeficientul de inteligență QI=200!) era expert în problemele de ereditate; americanul Reed, profesor de statistică la Baltimore etc.

La progresul statisticii și-au adus contribuția și unii statisticieni români, dintre care cităm: Ion Ionescu de la Brad, Dionisie Pop Marțian, Dr. Sabin Manuilă, Anton Golopenția, Dr. Petre Râmneanțu. S-ar părea că mînuirea datelor și formulelor statistice este rezervată numai anumitor persoane cu înclinații speciale pentru matematică. Într-adevăr, unele cercetări arată că există o anumită „specializare" a creierului pentru anumite funcții: emisfera dreaptă este „specializată" pentru activități artistice, de creație, de sinteză, ocupîndu-se mai mult de componenta emoțional-afectivă a comportamentului (urechea dreaptă

Scanned with OKEN Scanner

fiind urechea „muzicală") și se pare că îmbătrînește mai repede decît emisfera stîngă care este „specializată" pentru activități analitice, de logică, de gîndire abstractă, de matematică, fiind emisfera „intelectuală". Deci, după cum există artiști la care funcțiile emisferei drepte sînt foarte dezvoltate, tot așa sînt matematicieni la care funcțiile emisferei stîngi sînt foarte dezvoltate. Dar, statistic, la marea majoritate a oamenilor există un echilibru între dezvoltarea emisferelor cerebrale, deci statistica poate fi înțeleasă de majoritatea oamenilor, desigur cu puțină bunăvoință și efort intelectual.

Termenii acceptați în statistică sînt următorii:

Statistica este știința care se ocupă cu descrierea și analizarea numerică a fenomenelor de masă. Ea studiază latura cantitativă a fenomenelor, legile statistice manifestîndu-se sub formă de tendințe.

Populație, statistic, înseamnă o mulțime de elemente care aparțin aceleiași familii și fac obiectul cercetării statistice. Această denumire a rămas astfel, deoarece la început statistica se ocupa cu studiul populațiilor umane. Astăzi statistica se aplică și populațiilor animale, produselor agricole, comerciale etc., deci termenul nu mai este corect, motiv pentru care se mai întrebuințează termenul de „colectivitate", „mulțime" sau chiar „univers". Părțile componente se numesc „indivizi", „elemente", „unități", „cazuri", „termeni", „valori". După STAS 2631-72, populație sau colectivitate este o mulțime de obiecte sau fenomene calitativ omogene, formînd pe această bază ansambluri sau sisteme de același tip. Unitatea statistică este elementul component al populației statistice supus cercetării, elementele avînd unele caracteristici comune.

D.p.d.v. statistic prin *mulțime* se înțelege un ansamblu sau o colectivitate de elemente reunite după un criteriu sau o regulă, iar disciplina care studiază proprietățile generale ale mulțimilor, adică proprietățile independente de natura elementelor ce alcătuiesc mulțimea, este „teoria mulțimilor", disciplină matematică bine conturată. Cu ajutorul unor formule matematice aplicate la statistică, putem studia relațiile cantitative dintre fenomenele de masă, putînd generaliza la întreg (colectivitate, populație), ceea ce am găsit la particular (eșantion, lot, mostră).

Întrebuințăm acest termen și nu invers („statistică matematică"), deoarece nu facem un studiu statistic al

Scanned with OKEN Scanner

matematicii (întrebuințăm termenul de „statistică agricolă", „statistică economică" etc., deoarece efectuăm un studiu statistic aplicat la agricultură etc. — ori, în cazul de față nu facem un studiu statistic al matematicii, care să justifice termenul de „statistică matematică").

Statistica se ocupă cu elementele luate în masă, fără a putea ține seama de cazul izolat: individ sau moment. Claude Bernard, prin 1860 afirma în „Introducere în studiul medicinii experimentale" că medicul nu are ce face cu „legea numerelor mari", căci pe chirurg îl interesează de ce a murit pacientul după operație și nu îl interesează procentul de morți în situații asemănătoare. Aceasta se datorește faptului că pe vremea sa matematica nu era destul de bine cunoscută pentru a fi aplicată la studiul fenomenelor biologice. Și, într-adevăr, „într-o familie în care s-a produs un accident, nu servește la nimic să vorbim despre statistică".

În încheiere redăm un exemplu de aplicabilitate a metodelor statistice: Deeming, statistician american, a fost invitat în Japonia pentru a preda un curs de statistică, ce a fost urmat de 400 de ingineri din diferite ramuri. Acest curs a avut un asemenea succes în producție încît premiul anual pentru productivitate a fost denumit „premiul Deeming". Deeming a fost apoi angajat la o companie în anul 1979, unde prin cercetări statistice a uniformizat produsele, economisindu-se 1/2 milion de dolari, metodele statistice conducînd la creșterea productivității, angajații lucrînd cu mai multă inteligență și cu mai puțin efort fizic.

Scanned with OKEN Scanner

Capitolul 2

Noţiuni introductive în matematica statistică

Prelucrarea datelor statistice cu ajutorul unor formule matematice este necesară pentru că ne ajută să vedem şi să înţelegem fenomenele mai puţin evidente, mai puţin caracteristice. Din această cauză, considerăm că este necesară cunoaşterea sau reamintirea unor lucruri cu care ne vom întîlni pe parcurs, inclusiv noţiuni de informatică şi calculatoare.

2.1. Scrierea cifrelor

Pentru a evita greşelile sau confuziile, reamintim următoarele:

— Cifrele se pot scrie „în şir" sau „în coloană". Cifrele din şir se separă prin ; (punct şi virgulă), de ex.: 4; 5,5; 6; 8 etc., iar zecimalele prin [,] (virgulă). În literatura statistică mai veche se întrebuinţa de ex.: 2.150,25 (două mii cinci zeci, virgulă 25), iar în literatura anglo-americană se întrebuinţează de ex.: 35,000 (treizeci şi cinci de mii). De asemenea se mai întrebuinţează: .02 (zero virgulă zero doi), precum şi: 4.50 dolari (patru dolari şi cincizeci cenţi).

— Cifrele mari se scriu cu pauză între ordinul miilor, ex.: 35 000.

— Cifrele mari se pot scrie şi ca puteri, de ex.:

10 . . . . . 10^1
100 . . . . . 10^2
1 000 . . . . . 10^3
5 720 . . . . . $5{,}72 \cdot 10^3 = 0{,}572 \cdot 10^4$

Cifrele subunitare se pot scrie de asemenea ca puteri:

0,1 . . . . . 10^{-1}
0,01 . . . . . $10^{-2} = 1/100 =$ procent
0,001 . . . . . $10^{-3} = 1/1\,000 =$ promile $=$ g/kg $=$ ml/l

0,000 001 . . . 10^{-6}=1/1 000 000=ppm=mg/kg=0,1 ml/l
0,000 000 001 . 10^{-9}=1/1 000 000 000=ppb=gama/kg

— Gradul de precizie (a unei măsurători) se vede din zecimalele respective, de ex.: 10 g înseamnă 10 grame, dar 10,000 înseamnă 10 grame, cîntărit la balanța analitică, deci nici 1 mg în plus sau minus.

— Valorile proporționale se scriu sub forma: 10% (zece la sută), 2‰ (doi la mie) etc. Este de preferat ca la sfîrșitul măsurătorilor să se calculeze și să se exprime valori rotunjite, de ex. în loc de 1,62/81 să se calculeze și să se scrie: 2/100.

Se va ține seamă că în diferite țări exprimarea pentru unele cifre diferă, astfel:

CIFRA	ANGLIA, GERMANIA, ROMÂNIA	FRANȚA, S.U.A., U.R.S.S.
10^6	milion	milion
10^9	miliard	bilion
10^{12}	bilion	trilion
10^{15}	biliard	cvadrilion
10^{18}	trilion	cvintilion
10^{21}	triliard	sextilion
10^{24}	cvadrilion	septilion
10^{27}	cvadriliard	octilion

2.2. Șiruri (serii) de numere

Prin șir sau serie de numere înțelegem o succesiune de numere care au între ele o anumite corelație. Astfel, pot fi imaginate oricîte feluri de șiruri, dintre care cităm cîteva mai importante.

1. Șirul numerelor naturale: 1; 2; 3; 4; 5 etc. În acest caz, putem considera că ele vor crește în continuare în aceeași ordine naturală, putînd face și extrapolări, d.ex.:

— într-o experiență, temperatura crește la +100°C; +200°C etc., dar nu poate scădea la —100°C; —200°C;

—300°C etc., deoarece la —273,16°C este zero absolut = 0°Kelvin;

— într-un recipient presiunea poate crește la +1 atm; +2 atm etc., dar nu poate scădea la —1 atm; —2 atm etc.;

— înălțimea copiilor crește în funcție de vîrstă, totuși nu se poate ajunge la 3 . . . 4 m din cauza osificării cartilagiilor de creștere.

În *The Guiness Book of Records*, 1988 se citează că cel mai înalt bărbat a fost Robert Wadlow (1918—1940) din S.U.A. de 272 cm, cea mai înaltă femeie, Zeng Jimbian (1964—1982) din China de 247 cm; bărbatul cu greutatea cea mai mare a fost John Minoch (1941—1983) din S.U.A. 635 kg, iar ca femeie, Ida Maitland (1898—1932) din S.U.A. de 413 kg. Deci în extrapolări trebuie ținut seama și de logica faptelor.

2. Șirul logaritmic al numerelor naturale: 1; 10; 100; 1 000 etc.

3. Șirul lui Fibonacci: 1; 1; 2; 3; 5; 8; 13; 21 etc., cînd numărul următor se obține din sumarea celor două numere anterioare. Șirul lui Fibonacci se întîlnește în natură: conul pinului este format din 5 și 8 elemente, ananasul are 8 și 13 spire, pianul are 5 clape negre și 8 clape albe = 13 note etc. Relația dintre numerele din șirul lui Fibonacci (după cifra 3) este egală cu aprox. 1/1,6 proporție plăcută, cunoscută sub numele de „secțiunea de aur", ex. un tablou ce are formatul 1/1,6 are o „armonie" a dimensiunilor, proporție ce a fost utilizată în toate capodoperele antichității (înălțimea și lățimea templelor), în Renaștere (picturile lui Leonardo da Vinci), în epoca modernă (construcțiile lui Corbusier). „Numărul de aur" este dat de relația: $\frac{1+\sqrt{5}}{2} \approx \frac{5}{3} \approx \frac{8}{5}$ etc. $\approx 1{,}61$, iar „proporția de aur" este dată de relația $\frac{CB}{AC} = \frac{AC}{AB}$ (Fig. 3).

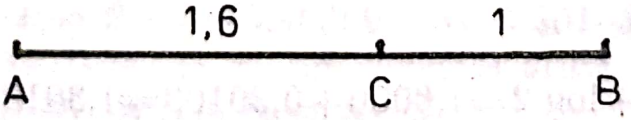


FIG. 3.

4. Șirul progresiv aritmetic, care are formula: $n_i = k + a(i-1)$, unde: i = numărul termenului din șir, k = constantă, a = rația progresiei.

Exemplul: 1; 3; 5; 7 etc., deci fiecare termen este egal cu termenul anterior plus o cantitate numită „rația progresiei".

5. Șirul progresiv geometric, care are formula: $n_i = k \cdot a^{(i-1)}$. Exemplu: 2; 4; 8; 16 etc., deci acest șir se deosebește de cel anterior prin faptul că cifra a se înmulțește cu constanta k. Acest șir are aplicații în biologie, de ex. putîndu-se calcula numărul unei populații după un anumit număr i de generații.

7. Șirul armonic, care provine din suma termenului anterior + inversul termenului natural următor, astfel:

1 provine din $\frac{1}{1}$

1,5 „ $1+\frac{1}{2}$

1,83 „ $1+\frac{1}{3}$ etc.

Acest șir de numere nu are aplicație în statistică.

2.3. Logaritmi

În anul 1614 matematicianul scoțian John Neper a inventat logaritmii naturali (numiți și logaritmi „neperieni") pe care Briggs i-a folosit pentru calcularea logaritmilor zecimali. Aplicînd logaritmii la unele calcule, acestea se simplifică foarte mult, iar altele nu s-ar putea efectua fără logaritmi. Reamintim modul de lucru cu logaritmii:

— Înmulțirea se transformă în adunare, de ex.: $10 \cdot 2 = x$

$$\log x = \log 10 + \log 2 = 1{,}0000 + 0{,}30103 = 1{,}30103$$

Antilogaritmul mantisei 30103=2. Avînd caracteristica „1", înseamnă că avem 2 numere întregi, deci x=20.

— Împărțirea se transformă în scădere, de ex. $10 : 2 = x$

$\log x = \log 10 - \log 2 = 1{,}0000 - 0{,}30103 = 0{,}69897$

Antilogaritmul mantisei 69897=5. Avînd caracteristica „0“ înseamnă că vom avea 1 număr întreg, deci $x=5$.

— Ridicarea la putere se transformă în înmulțire, de ex. $8^{4/2} = x$

$$\log x = \frac{4 \cdot \log 8}{2} = \frac{4 \cdot 0{,}90309}{2} = 1{,}80618$$

Antilogaritmul mantisei 80618=64. Avînd caracteristica „1“, vom avea 2 numere întregi, deci $x=64$.

Problema se mai poate rezolva astfel: $8^2=64$. Dar se mai poate rezolva și astfel: $8^{4/2}\sqrt[2]{8^4} = = \sqrt[2]{4096} = 64$.

Aplicînd logaritmii la putere negativă, se procedează astfel:

$25^{-2} = x$, de unde: $\log x = -2 \cdot \log 25 = -2 \cdot 1{,}39794 = = -2{,}79588$. Deoarece rezultatul logaritmării este negativ (și în tabelele de logaritmi cifrele sînt calculate numai pentru valori pozitive), se determină cologaritmul astfel: la caracteristică se adaugă „—1“, deci vom avea: $-2 + +(-1) = -3$, iar mantisa se scade din cifra 9999 10 astfel:

9999 10—
7959 8

2041 2, deci cologaritmul este —3,20412. Antilogaritmul mantisei 20412=160. Deoarece caracteristica este „—3“, vom avea 3 zero-uri, deci $x=0{,}00160$.

Problema se mai poate rezolva și astfel: $\frac{1}{25^2} = \frac{1}{625} = 0{,}0016$.

— Extragerea unei rădăcini se transformă în împărțire, de ex. $\sqrt[3]{8} = x$:

$\log x = \log 8 : 3 = 0{,}90309 : 3 = 0{,}30103$. Antilogaritmul mantisei 30103=2 și avînd caracteristica „0“ înseamnă că avem 1 număr întreg, deci $x=2$.

Revenim la șirul geometric progresiv, rezolvat cu logaritmi: Shirdan, sultan în India, după legendă, a vrut să răsplătească pe inventatorul jocului de șah astfel: 1 bob

de grîu pentru pătrățica 1, 2 pentru cea de-a doua, 4 pentru cea de-a treia, 8 pentru cea de-a patra, 16 pentru cea de-a cincea etc., lucru ușor la prima vedere. Dar făcîndu-se calculele s-a găsit că revenea un număr enorm de boabe. Calculul se face astfel: $\sum_{i=1}^{m} n_i = a^m - 1$, unde:

$\sum_{i=1}^{m} n_i$ = suma termenilor de la $i=1$ pînă la termenul m

$a=2$, deoarece se dublează de fiecare dată numărul de boabe

$m=64$, după numărul pătrățelelor de la șah.

Deci avem:

$\Sigma n_{64} = 2^{64} - 1$. Logaritmînd avem: $64 \cdot \log 2 = 64 \cdot 0{,}30103 = 19{,}26592$ (nu am mai introdus în calcul pe „—1“, deoarece la suma totală nu are importanță). Antilogaritmul mantisei 26592=1845. Caracteristica fiind „19“, vom avea 20 numere întregi la rezultat, deci $\Sigma n_{64} = 18{,}45 \cdot 10^{18}$ boabe de grîu. Știind că 1 kg grîu are aprox. 250 000 boabe ($=2{,}5 \cdot 10^5$) avem $(18{,}45 \cdot 10^{18}) : (2{,}5 \cdot 10^5) = 7{,}38 \cdot 10^{13}$ kg grîu $= 73{,}8 \cdot 10^{12} = 73{,}8$ bilioane tone grîu, deci cîteva recolte anuale de pe întreg Pămîntul!

Corelația logaritmilor cu unele funcții biologice este evidentă în multe situații, dintre care cităm: ca o măsură de prevedere împotriva șocurilor optice, acustice etc., senzația fiziologică nu crește decît aritmetic, deși acțiunea fizică are creștere logaritmică (în baza 10), de ex.:

1 bec de 10 lucși dă senzația de lumină „1“,

1 bec de 100 lucși dă senzația de lumină „2“, adică apare pentru ochi de 2 ori mai puternică, nu de 10 ori,

1 bec de 1 000 lucși dă senzația de lumină „3“, adică apare pentru ochi de 3 ori mai puternică, nu de 100 de ori. Aceste cifre sînt în concordanță cu faptul că logaritmul lui 100=2, logaritmul lui 1 000=3 etc., toate acestea în concordanță cu faptul că legea lui Fechner stabilește că senzația fiziologică variază cu logaritmul (în baza 10) al senzației fizice, ceea ce confirmă legea lui Weber, care stabilește că senzațiile cresc în funcție de proporțiile unităților, nu de numărul acestora.

Redăm un tabel prescurtat cu logaritmi.

Scanned with OKEN Scanner

N	LOG (BAZA=2)	LN NATURAL (BAZA=e=2,71)	LG ZECIMAL (BAZA=10)
1	0	0	0
2	1,0000	0,69315	0,30103
3	1,5850	1,09861	0,47712
4	2,0000	1,38624	0,60206
5	2,3219	1,60944	0,69897
6	2,5850	1,79176	0,77815
7	2,8074	1,94591	0,84510
8	3,0000	2,07944	0,90309
9	3,1699	2,19722	0,95424
10	3,3219	2,30259	1,00000
20	4,3219	2,99573	1,30103
25	4,6439	3,21888	1,39794
32	5,0000	3,46573	1,50515
64	6,0000	4,15888	1,80618
100	6,64	4,60517	2,00000

2.4. Maşini de calcul

Cele mai vechi mărturii ale unor calcule sînt tăbliţele de lut ale vechilor sumerieni (3100—3000 ani î.e.n.). Calculele se făceau cu pietricele, cu bucăţi de lemn (în sistemul dual), apoi cu degetele (în sistemul zecimal). Civilizaţia a păstrat urme ale diferitelor sisteme de numărare, de ex. duzina (12 bucăţi), gros-ul (12 duzini), 12 ore ale zilei, 12 luni ale anului etc. La francezi, din sistemul „20“ a mai rămas 80 (care se spune: quatre-vingt=4 ori 20) şi 90 (care se spune: quatre-vingt-dix=4 ori 20 şi 10).

Rigla de calcul

Una din primele „maşini“ de calcul este rigla de calcul, construită după inventarea logaritmilor. Ea a fost construită de englezul Seth Patrige, în Franţa fiind introdusă de Jonard. Pe riglă sînt înscrise numerele, patratele, rădăcinile patrate etc. Există rigle pentru calcule forestiere, pentru construcţii, calcule electrice etc.

Scanned with OKEN Scanner

Maşini mecanice de calcul

Aceste maşini sînt foarte variate: cu bile (abac-uri), cu cilindri, cu discuri, cu roţi dinţate etc. Matematicianul Blaise Pascal, la vîrsta de 19 ani, a construit prima maşină de adunat şi scăzut în 1642. În 1761 matematicianul german Leibniz a construit prima maşină pentru cele patru operaţii, îmbunătăţită în 1779 de Matheus Hahn. În 1833 englezul Charles Babage, profesor la Universitatea din Cambridge, concepe principiul unei maşini analitice, iar în 1858 englezul Xavier Thomas construieşte maşini de calculat în serie. În 1889 austriacul Hermann Holerith pune la punct un sistem de codaj prin cartele perforate (inventate de francezul Falcon şi întrebuinţate la războaiele de ţesut tip Jaquard, apoi utilizate în recensămînt). În 1888 americanul Borrugs a brevetat o maşină de adunat cu tastatură şi imprimare, care în 1920 se dotează cu motor electric pentru mărirea vitezei de calcul şi se completează cu relee electro-magnetice.

2.5. Maşini electronice de calcul

În 1942—46 a fost construită prima maşină electronică de calcul de către Mauxhley şi Eckert în S.U.A., care avea 18 000 de lămpi de radio, ocupa 160 mp şi cîntărea 30 de tone. În Anglia, sub direcţia lui Wilkes se construieşte primul calculator, de 10 ori mai rapid, cu 3 000 de lămpi de radio. În 1951 Biroul american de recensămînt achiziţionează prima maşină de calcul modernă pentru întrebuinţări civile, iar în 1953 uzina General Electric achiziţionează prima maşină electronică de calcul pentru întrebuinţarea în sector particular. La noi, primele maşini electronice de calcul au fost construite în 1950 la Institutul de fizică atomică Bucureşti. Miniaturizarea componentelor electronice a făcut ca astfel de maşini să poată fi acum purtate în buzunar, numite deci şi „calculatoare de buzunar", avînd o serie de funcţii şi efectuînd calculele cu mare viteză. Aceste maşini se pot clasifica astfel:

generaţia I: 1947, cu lămpi de radio şi 5 000 operaţii/s. Calculele se făceau succesiv.

Scanned with OKEN Scanner

generația II: 1958 cu tranzistori și circuite imprimate. Calculele se făceau simultan (20 000 operații/s).

generația III: 1979 cu tranzistori și circuite integrate, cu 2 milioane de operații/s.

generația IV: 1979 cu circuite de înaltă integrare, cu zeci de milioane operații/s.

generația V: 1990, cu supraconductori, care nu calculează, ci „inferează" (deduce) cu viteza de 1 miliard inferențe/s.

Dezvoltarea acestor calculatoare se datorește apariției circuitelor integrate, evoluția lor fiind următoarea:

SSI=Smale Scale Integration (integrare în măsură redusă)

MSI=Medium Large Scale Integration (integrare pe scară foarte largă)

VLSI=Verry Large Scale Integration (integrare pe scară foarte mare)

ULSI=Ultra Large Scale Integration (integrare extrem de mare), iar după 1990 se prevede apariția unui „biocip", adică un circuit integrat pe bază de materie organică. Aceste mașini, denumite la început „computere", „creier electronic", „ordinator" etc., astăzi se numesc la noi „calculator". Ele efectuează calculele în sistem binar (fiind mai ușor d.p.d.v. constructiv), rezultatele fiind apoi „traduse" în sistem zecimal. La calculatoarele centrale pot fi conectate mai multe „terminale" (analog cu legarea posturilor telefonice la centrala telefonică) de unde se pot obține informații, calcule etc., deoarece în memoria acestora se pot stoca foarte multe informații. Informațiile ce se introduc în calculator sînt formulate într-un limbaj convențional pentru a fi „înțelese" de calculator. Astfel de limbaje sînt:

FORTRAN (FORmula TRANslator) apărut în 1959, utilizat în industrie, ALGOL (ALGOritmic Langage), apărut în 1960, utilizat mai ales în cercetarea științifică, COBOL (Common Business Oriented Langage) utilizat mai ales în contabilitate, SOAP (Symbolic Optimal Assembly Program) etc. Mai nou, există și limbaje „convenționale" ca APL, BASIC (Beginer's All purpose Symbolic Instruction Code) apărut în 1964, care se pot implementa în microprocesoare prin construcție, ceea ce reduce din timpul

Scanned with OKEN Scanner

de programare. Unele aparate de laborator sînt atît de perfecționate prin adăugarea de calculatoare electronice, încît într-un prospect se spune: Dv. alegeți felul cercetării și calculatorul, grație unui program inteligent, întreține cu dv. un „dialog"în timpul cercetării, pînă la obținerea rezultatului. Rezultatul se poate stoca în memorie, se poate exprima cifric sau grafic după dorință, evitîndu-se greșelile de transcriere.

Allan Mathison Turing (1912—1954) matematician englez, a prezis în 1950 că în anul 2000 se va putea susține o conversație cu un calculator, prin telescriptor, omul nefiind dumirit dacă îi răspunde un om sau o mașină. Astăzi, această anticipare este depășită. Calculatoarele perfecționate au în program „SOA", adică un sistem de operare algebrică ce execută operațiile matematice (pe lîngă alte operații), după un anumit algoritm (succesiune de operații care conduc la un anumit rezultat), în următoarea ordine:

1. funcții cu variabile singure (logaritmi, sin, cos etc.)
2. funcții exponențiale și funcții putere
3. înmulțiri, împărțiri
4. adunări, scăderi
5. rezolvă parantezele și le memorează.

De fapt, și fără calculator aceste operații matematice trebuie rezolvate în aceeași ordine pentru a obține rezultate corecte, și anume:

1. operații de ordinul 3: puteri, rădăcini, funcții, logaritmi
2. operații de ordinul 2: înmulțiri, împărțiri
3. operații de ordinul 1: adunări, scăderi

Dezvoltarea industriei calculatoarelor a făcut ca unele prognoze să prevadă că în viitor calculatoarele vor fi mai mici, mai ieftine, mai rapide și că vor „învăța" din calculele anterioare.

Inventarea microprocesorului (în 1971) este un moment istoric în stocarea datelor (informațiilor) într-un volum mic. Aceste microprocesoare se întrebuințează la automatele din industrie, în telecomunicații etc.

Scanned with OKEN Scanner

2.6. Noțiuni de informatică

Prin *informație* înțelegem fiecare din elementele conținute în semnificația unui simbol sau a unui grup de simboluri (comunicare verbală, text scris etc.) cu ajutorul cărora se comunică date relative la evenimente, stări, situații, acțiuni etc.

Teoria informației este știința care se ocupă d.p.d.v. probabilistic cu legile cantitative privind obținerea, transmiterea, prelucrarea și stocarea informațiilor, fără a se interesa de felul acestora: optice, acustice, electrice, termice, mecanice etc. În aceasta se include și studiul componentelor „hardware" (ansamblul dispozitivelor sau componentelor electrice, mecanice, magnetice etc.) ale calculatoarelor și „software" (modul de programare care cuprinde limbajele utilizate, algoritmii etc.), precum și „firware" (programarea inițială, fixată în structura și memoria permanentă a calculatorului). S-ar putea face o paralelă: în creier sînt o serie de structuri nervoase (hardware) cu o serie de instincte (firmware) precum și o serie de aptitudini sau cunoștințe cîștigate (software).

Unitatea de informație este bit-ul (BInar digiT information), care este rezultatul binar al unei informații net deosebite, ex. sus—jos; aprins—stins; viu—mort, închis—deschis etc. Organismul uman emite prin vorbire, mimică, gesturi etc. aprox. 10^7 biți/s și reține pentru prelucrare (ca o măsură de autoapărare) doar 10^2 biți, din care se rețin în memorie doar 1—2 biți/s. Pînă acum nu se cunosc mecanismele prin care se face această selecție a informațiilor (una din definițiile inteligenței este: capacitatea de selecție a informațiilor). O cifră (din 10) ne dă o informație de 3,32 biți (conform logaritmului în baza 2), o literă (din cele 25 ale alfabetului) ne dă 4,64 biți, iar o pagină dintr-o carte conține cîteva mii de biți.

Unitățile de informație sînt următoarele:

bit=unitatea, care nu are submultipli
kilobit=1 000 biți=kb
megabit=1 milion biți=Mb
octet=8 biți=0 (care se mai numește și „byte")
kilo-octet=1 000 octeți=ko (exact: 2^{10}=1 024 octeți=8 192 biți)

Scanned with OKEN Scanner

megaoctet=1 milion octeți=Mo (exact: 2^{20} octeți=8 388 608 octeți).

Am redat aceste date, deoarece în unele prospecte sînt enumerate calitățile calculatoarelor, între care și capacitatea de memorare în biți. Unele calculatoare au mii de Mo,

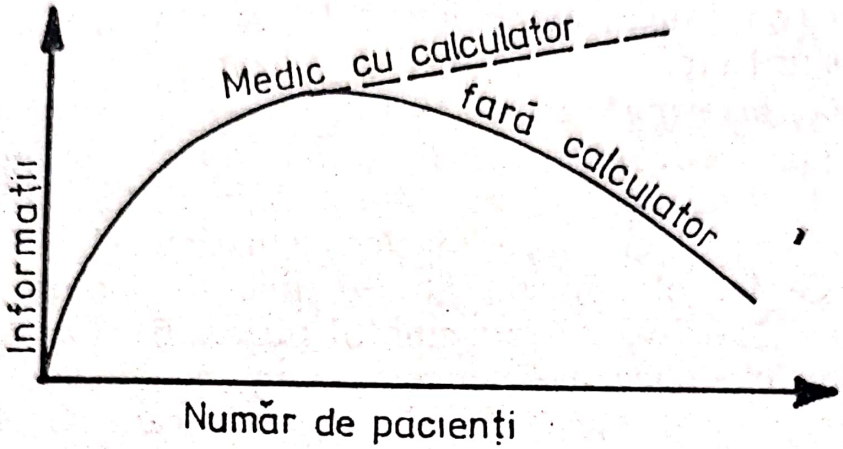


FIG. 4.

memorie ce se poate extinde aproape la infinit prin adăugare de discuri magnetice, benzi etc., calculatoarele centrale putînd furniza informații foarte diverse (dacă acestea au fost introduse în memorie).

Datorită memoriei lor, calculatoarele vor fi strict necesare în diverse domenii de activitate. Un astfel de exemplu este redat în fig. 4 unde se vede că posibilitățile de supraveghere a bolnavilor de către medic scad dacă numărul pacienților crește, dar dacă este „ajutat" de calculator, numărul pacienților poate crește. O prognoză prevede că pînă în 1990 toate activitățile economice vor fi rezolvate cu calculatorul, chiar și cheltuielile casnice ale gospodinelor. Perfecționarea în domeniul informaticii se vede și din următorul exemplu: în 1963 prin rețeaua de telex (în Franța) se puteau transmite 50 „baud" (=50 biți/s); astăzi se pot transmite prin rețele speciale 2 milioane de baud („baud" este denumirea dată în onoarea inginerului francez Baudot Emil).

Utilizarea calculatoarelor „de buzunar"

Întrucît există multe modele, trebuie studiată cu atenție cartea cu instrucțiunile fabricii pentru a exploata toate posibilitățile de lucru ale calculatorului, în funcție de structura sa (hardware), de programarea introdusă în me-

moria sa (firmware). În principiu, trebuie ținut seama de următoarele:

— alimentarea corectă cu energie electrică (baterii, acumulatori, lumină suficientă la cele cu fotocelule). La alimentare necorespunzătoare apar greșeli de calcul. Verificarea se face, de ex. 1111^2, rezultatul trebuind să fie 1234321;

— evitarea temperaturilor prea scăzute sau prea ridicate;

— evitarea vaporilor de apă, acizi, solvenți organici;

— evitarea zguduirilor puternice (cădere pe podea);

— în timpul neutilizării, comutatorul se închide, iar la nefolosire îndelungată se scot bateriile din calculator;

— calculatoarele cu cristale lichide consumă mai puțin curent decît cele cu afișaj LED (Light Emisiv Diode).

Unele calculatoare sînt în combinație cu un ceas ale cărui cifre dispar în timpul calculelor, ceasul urmîndu-și funcționarea în mod independent de calculator. Calculatoarele simple servesc de obicei pentru cele patru operații de bază, existînd însă și clape pentru rădăcină patrată, procentaj, schimbarea semnului + în — (și invers), memorizare și anularea memorizării. Cele mai frecvente notații la calculatoarele de buzunar sînt următoarele:

Ein=On=deschis=funcționează

Aus=Of=afară=închis=se întrerupe funcționarea

M=introducerea în memorie

MR; MRC=memory recall=readucerea din memorie

MC=Memory cleared=ștergerea din memorie

C=Correction=corectare=ștergerea ultimei cifre introduse

AC=All Cleared=șterge totul

CE=Cleare Error=șterge eroarea

Reset=șterge, anulează totul

Del=Deletion=ștergere

Mai trebuie menționate pentru calculatoarele mari:

ROM=Read Only Memory=memorie „moartă"=memorie definitivă

RAM=Random Acces Memory=memorie „vie"=memorie nepermanentă

PROM=Programable Read Only Memory=memorie permanentă, din fabricație

Scanned with OKEN Scanner

moria sa (firmware). În principiu, trebuie ținut seama de următoarele:

— alimentarea corectă cu energie electrică (baterii, acumulatori, lumină suficientă la cele cu fotocelule). La alimentare necorespunzătoare apar greșeli de calcul. Verificarea se face, de ex. 1111^2, rezultatul trebuind să fie 1234321;

— evitarea temperaturilor prea scăzute sau prea ridicate;

— evitarea vaporilor de apă, acizi, solvenți organici;

— evitarea zguduirilor puternice (cădere pe podea);

— în timpul neutilizării, comutatorul se închide, iar la nefolosire îndelungată se scot bateriile din calculator;

— calculatoarele cu cristale lichide consumă mai puțin curent decît cele cu afișaj LED (Light Emisiv Diode).

Unele calculatoare sînt în combinație cu un ceas ale cărui cifre dispar în timpul calculelor, ceasul urmîndu-și funcționarea în mod independent de calculator. Calculatoarele simple servesc de obicei pentru cele patru operații de bază, existînd însă și clape pentru rădăcină patrată, procentaj, schimbarea semnului + în — (și invers), memorizare și anularea memorizării. Cele mai frecvente notații la calculatoarele de buzunar sînt următoarele:

Ein=On=deschis=funcționează

Aus=Of=afară=închis=se întrerupe funcționarea

M=introducerea în memorie

MR; MRC=memory recall=readucerea din memorie

MC=Memory cleared=ștergerea din memorie

C=Correction=corectare=ștergerea ultimei cifre introduse

AC=All Cleared=șterge totul

CE=Cleare Error=șterge eroarea

Reset=șterge, anulează totul

Del=Deletion=ștergere

Mai trebuie menționate pentru calculatoarele mari:

ROM=Read Only Memory=memorie „moartă"=memorie definitivă

RAM=Random Acces Memory=memorie „vie"=memorie nepermanentă

PROM=Programable Read Only Memory=memorie permanentă, din fabricație

Scanned with OKEN Scanner

EPROM=Erassable PROM=memorie reprogramabilă.

Cu cît aceste valori sînt mai mari, cu atît capacitatea de lucru este mai mare. La calculatoarele „științifice", mai complicate, se pot rezolva probleme de trigonometrie, statistică, de ecuații de diferite grade etc. și în funcție de microprocesorul respectiv, se pot utiliza la jocuri pe ecran de TV (tenis, fotbal etc.).

Din datele prezentate rezultă că statistica nu este o achiziție recentă. De asemenea se vede că statistica (și calculatoarele) sînt rezultatul comun al inteligenței și strădaniei multor generații de oameni, din diferite epoci, din diferite țări, de diferite profesiuni. Progresul cel mai mare s-a făcut în secolul nostru prin inventarea calculatoarelor. Previziuni? Sînt neprevizibile!

Scanned with OKEN Scanner

Capitolul 3

Calcule statistice

3.1. Medii

Atenuarea datelor, cifrelor, accidental prea mari sau prea mici dintr-o serie de determinări, măsurători, analize etc. se face prin calcularea mediei acestor valori, făcîndu-se o compensare a abaterilor individuale. Acest calcul ne arată tendința fenomenelor, media statistică fiind o valoare ce sintetizează într-o singură expresie numerică toate valorile din seria determinărilor. Termenii seriei diferă de medie, deoarece au fost influențați de diferiți factori.

3.1.1. Media aritmetică

Media aritmetică simplă

Exprimă nivelul mediu, anihilînd abaterile individuale, netipice. Ea este cuprinsă între valoarea cea mai mare și cea mai mică. Fiind simplu de calculat, este metoda cea mai utilizată. Are dezavantajul că este sensibilă la valorile extreme, iar dacă termenii sînt prea „împrăștiați", tinde să devină o valoare nereprezentativă. Media aritmetică este o valoare lipsită de conținut dacă elementele sînt net deosebite d.p.d.v. calitativ, caz în care este mai util a se stabili medii parțiale pentru fiecare tip de colectivitate. Media aritmetică este definită ca suma valorilor ce alcătuiesc colectivitatea, împărțită la numărul loir. Deci este un raport între suma termenilor și numărul lor. Formula sa este următoarea:

$$\bar{x}_{as} = \frac{x_1 + x_2 + \ldots x_n}{n} = \frac{\Sigma x_i}{n}, \text{ unde:}$$

x_i = valorile individuale

n = numărul valorilor individuale.

Scanned with OKEN Scanner

Capitolul 3

Calcule statistice

3.1. Medii

Atenuarea datelor, cifrelor, accidental prea mari sau prea mici dintr-o serie de determinări, măsurători, analize etc. se face prin calcularea mediei acestor valori, făcîndu-se o compensare a abaterilor individuale. Acest calcul ne arată tendința fenomenelor, media statistică fiind o valoare ce sintetizează într-o singură expresie numerică toate valorile din seria determinărilor. Termenii seriei diferă de medie, deoarece au fost influențați de diferiți factori.

3.1.1. Media aritmetică

Media aritmetică simplă

Exprimă nivelul mediu, anihilînd abaterile individuale, netipice. Ea este cuprinsă între valoarea cea mai mare și cea mai mică. Fiind simplu de calculat, este metoda cea mai utilizată. Are dezavantajul că este sensibilă la valorile extreme, iar dacă termenii sînt prea „împrăștiați", tinde să devină o valoare nereprezentativă. Media aritmetică este o valoare lipsită de conținut dacă elementele sînt net deosebite d.p.d.v. calitativ, caz în care este mai util a se stabili medii parțiale pentru fiecare tip de colectivitate. Media aritmetică este definită ca suma valorilor ce alcătuiesc colectivitatea, împărțită la numărul loir. Deci este un raport între suma termenilor și numărul lor. Formula sa este următoarea:

$$\bar{x}_{as} = \frac{x_1 + x_2 + \ldots x_n}{n} = \frac{\Sigma x_i}{n}, \text{ unde:}$$

x_i = valorile individuale

n = numărul valorilor individuale.

Scanned with OKEN Scanner

Exemplul 1:

Fie valorile: 49; 51. Care este media acestor valori?

$$\bar{x} = \frac{49+51}{2} = 50$$

Exemplul 2:

Fie valorile: 0; 100. Care este media acestor valori?

$$\bar{x}_{as} = \frac{0+100}{2} = 50$$

Deci media aritmetică este aceeaşi ca în exemplul 1, unde este „normal" ca media să fie 50, aceasta fiind o valoare „la mijlocul" dintre cele 2 cifre, pe cîtă vreme la exemplul 2 rezultatul este corect d.p.d.v. matematic, dar nu e corect totdeauna d.p.d.v. statistic, „distanţa" dintre cifrele mediate fiind foarte mare, adică valorile sînt foarte „împrăştiate".

Exemplul 3:

Fie valorile: 1; 2; 100. Care este media acestor valori?

$$\bar{x}_{as} = \frac{1+2+100}{3} = \frac{103}{3} = 34{,}3$$

Exemplul 4:

Fie valorile: 1+99+100. Care este media acestor valori?

$$\bar{x}_{as} = \frac{1+90+100}{3} = \frac{200}{3} = 66{,}6$$

Deci, după cum se vede, media aritmetică simplă este sensibilă la valorile extreme. Din acest motiv, în unele cazuri se întrebuinţează o altă medie, în funcţie de problema respectivă.

Exemplul 5:

Într-un judeţ sînt trei spitale, care au avut activitate diferită în decursul unui an. Care este activitatea lor medie? Situaţia a fost următoarea:

spitalul A cu 100 paturi a avut 1 000 bolnavi/an;

Scanned with OKEN Scanner

spitalul B cu 150 paturi a avut 2 000 bolnavi/an;
spitalul C cu 50 paturi a avut 200 bolnavi/an.

Calculul se face astfel:

$$\bar{x}_{as\ A} = \frac{1\,000}{100} = 10 \quad \text{bolnavi/pat/an}$$

$$\bar{x}_{as\ B} = \frac{2\,000}{150} = 13{,}3 \quad \text{bolnavi/pat/an}$$

$$\bar{x}_{as\ C} = \frac{200}{50} = 4 \quad \text{bolnavi/pat/an}$$

Valorile rezultate sînt „medii aritmetice parțiale" sau „medii pe grupe", pe care le putem utiliza pentru intercompararea utilizării paturilor celor trei spitale. Calculînd media activității spitalelor, $\bar{x}_{as\ A,\ B,\ C} = \frac{10+13{,}3+4}{3} = \frac{27{,}3}{3} = 9{,}1$ bolnavi/pat/an.

Această cifră este o „medie a mediilor", care nu reflectă realitatea, deoarece s-a dat aceeași importanță fiecărei „medii parțiale" indiferent de numărul elementelor din care au provenit numărul paturilor și numărul bolnavilor. Din această cauză (și din cauză că valorile sînt destul de disparate) trebuie utilizată altă metodă de calcul. Acest exemplu îl vom relua la fiecare calcul al diferitelor medii și în final la interpretarea rezultatelor.

Exemplul 6:

Fie următoarele temperaturi ale aerului într-o zi: ora 01=17°C; ora 07=18°; ora 13=25°; ora 19=20°. Care este media temperaturii?

$$\bar{x}_{as} = \frac{17+18+25+20}{4} = \frac{80}{4} = 20°C.$$

S-ar putea obiecta că temperatura de 25° a durat abia 2 ore, așa că totalul caloriilor primite de la soare este mai mic decît dacă ar fi durat de ex. 6 ore, deci $\bar{x}_{as}$ nu reprezintă realitatea. Răspunsul este următorul: este internațional acceptat ca măsurarea temperaturii aerului să se facă la orele menționate, făcîndu-se apoi media. Pentru alte determinări, se utilizează alte aparate (nu termometre); de ex. pentru cantitatea de căldură se utilizează

Scanned with OKEN Scanner

actinometre înregistratoare, care integrează totalul caloriilor, pentru durata însoririi se utilizează heliograful etc.

Exemplul 7:

Într-un recipient am avut următoarele valori: —2°C; 0°C; +2°C. Care este media temperaturilor?

$$\bar{x}_{as} = \frac{(-2)+0+2}{3} = \frac{0}{3} = 0°C$$

Deci, poate exista și media „zero".

Exemplul 8:

Temperatura unui bolnav este: 37° la ora 7; 38° la ora 12 și 38° la ora 18. Care este media temperaturii?

Matematic, $\bar{x}_{as}$=38°C, dar acest rezultat este fals, căci trebuie luat în considerare fiecare „rezultat parțial" și mai ales temperatura cea mai mare. Deci în acest caz, nu se face nici o medie.

Exemplul 9:

O unitate agricolă are 5 secerători, 10 semănători, 10 tractoare. Care este media mașinilor agricole?

Matematic $\bar{x}_{as}$= 8,3, dar această valoare nu reprezintă nimic, deoarece mașinile respective nu au nici o calitate comună, deci nu se face nici o medie.

În concluzie, pentru a calcula o medie, oricare ar fi ea, este nevoie de o anumită omogenitate a colectivității (d.p.d.v. al calității sau al proprietăților). Dacă termenii sînt prea „împrăștiați" (deci diferă mult d.p.d.v. cantitativ) este de preferat a se face „medii parțiale" pentru fiecare tip cantitativ al colectivității.

Media aritmetică ponderată

Se utilizează cînd valorile individuale x_i sînt înregistrate de mai multe ori, deci cînd avem serii de distribuții, adică avem frecvențe f_i neegale ale valorilor individuale x_i. Numărul care arată de cîte ori se repetă fiecare valoare (f_i) este „ponderea" valorii respective. Formula este următoarea:

$$\bar{x}_{ap} = \frac{(x_1 \cdot f_1)+(x_2 \cdot f_2)+ \ldots (x_n \cdot f_n)}{f_1+f_2+ \ldots f_n} = \frac{\Sigma (x_i \cdot f_i)}{\Sigma f_i}$$

Scanned with OKEN Scanner

Exemplul 1:

Reluînd exemplul 5 de la $\bar{x}_{as}$, problema se rezolvă astfel:

$$\bar{x}_{ap} = \frac{(100 \cdot 10)+(150 \cdot 13{,}3)+(50 \cdot 4)}{100+150+50} = \frac{3\,200}{300} =$$

$$= 10{,}66 \text{ bolnavi/pat/an.}$$

Dacă am fi efectuat calculul astfel:

$$\bar{x}_{ap} = \frac{(100 \cdot 10)+(150 \cdot 13{,}3)+(50 \cdot 4)}{10+13{,}3+4} = \frac{3\,200}{27{,}3} = 117{,}2$$

rezultat care nu are nici o semnificație. O astfel de rezolvare ar fi avut rost dacă problema ar fi fost următoarea: 10 persoane au cîștigat la loterie cîte 100 lei; 13 persoane au cîștigat cîte 150 lei; 4 persoane au cîștigat cîte 50 lei. Care este media cîștigului pe persoană? În acest caz ponderea ar fi fost dată de 10, 13 și 4 persoane, aceste cifre găsindu-se atît la numărător cît și la numitor. Deci ponderea ar fi fost dată de numărul persoanelor.

Exemplul 2:

10 persoane au avut tensiune arterială 130 mm Hg, iar 20 persoane au avut 145 mm Hg. Care este media tensiunii arteriale?

$$\bar{x}_{ap} = \frac{(10 \cdot 130)+(20 \cdot 145)}{10+20} = \frac{4\,200}{30} = 140 \text{ mm Hg/persoană}$$

Media aritmetică procentuală

Se utilizează cînd valorile sînt date procentual. Formula este:

$$\bar{x}_{a\%} = \frac{(p_1 \cdot x_1)+(p_2 \cdot x_2)+ \cdot \cdot (p_n \cdot x_n)}{100} = \frac{\Sigma (p \cdot x)}{100}, \text{ unde } p = \text{procentul.}$$

Exemplul 1:

10 bolnavi au ocupat		33,3%	din paturi la spitalul	A
13,3	„	50	„	B
4	„	16,6	„	C

Care este media bolnavilor?

$$\bar{x}_{a\%} = \frac{(10 \cdot 33{,}3)+(13{,}3 \cdot 50)+(4 \cdot 16{,}6)}{100} = \frac{1\,064}{100} =$$

$$= 10{,}64 \text{ bolnavi/an/pat.}$$

Scanned with OKEN Scanner

3.1.2. Media geometrică

Media geometrică simplă

Este mai puțin sensibilă la valorile extreme decît celelalte medii, deci se întrebuințează cînd dorim să atenuăm divergențele mari dintr-o serie de determinări cu frecvențe egale, fiind după o expresie „cea mai exactă medie". Se utilizează cînd valorile au o evoluție (de creștere sau scădere) permanentă, neîntreruptă, cu o rație din ce în ce mai mare, termenii fiind legați între ei printr-o relație de produs. De asemenea se mai întrebuințează cînd vrem să dăm o importanță mai mare termenilor mai mici, în valoare absolută, sau cînd diferențele dintre termeni sînt foarte mari. Are dezavantajul că nu se poate întrebuința cînd avem valori nule sau negative.

Datorită faptului că se calculează mai ușor cu ajutorul logaritmilor, se mai numește și „medie logaritmică". Putîndu-se utiliza și la calcularea ritmului (de creștere sau scădere), se mai numește și „medie de ritm". Este egală sau mai mică decît media aritmetică, dar mai mare decît media armonică. În rezumat, se întrebuințează cînd:

— seria este dinamică, fiind relații de creștere sau scădere,

— termenii au variații mari,

— distribuția are caracter pronunțat de asimetrie. Formula sa este următoarea:

$$\bar{x}_{geom\ s} = \sqrt[n]{x_1 \cdot x_2 \ldots x_n} = \sqrt[n]{\pi x_i},$$

unde π = „produsul dintre". Deci $\bar{x}_{geom\ s}$ este radicalul de ordin n din produsul valorilor individuale x_i de mărime pozitivă.

Se utilizează de ex. la aflarea concentrației medii a două gaze ce acționează într-o cameră de reacție, deoarece $\bar{c} = \sqrt{c_1 \cdot c_2}$ la egalitatea celor două componente, ceea ce are importanță în stabilirea timpului de reacție. În cazul particular cînd $n=2$, $\bar{x}_{geom\ s}$ a valorilor m_1 și m_2 se numește și „medie proporțională" între aceste valori căci $\bar{x}_{geom\ s} = \sqrt{m_1 \cdot m_2}$ se mai poate scrie: $\frac{m_1}{\bar{x}_{geom}} = \frac{\bar{x}_{geom}}{m_2}$.

Exemplul 1:

Fie valorile: 0; 4. Să se determine $\bar{x}_{as}$ și $\bar{x}_{geom\ s}$.

$$\bar{x}_{as}=\frac{0+4}{2}=2$$

$$\bar{x}_{geom\ s}=\sqrt{0.4}=\sqrt{0}=0.$$

Deci $\bar{x}_{geom\ s}$ nu se poate utiliza cînd avem valori nule, spre deosebire de $\bar{x}_{as}$, ce se poate utiliza.

Exemplul 2:

Fie valorile: 4; 6; 7; 10. Să se determine $\bar{x}_{as}$ și $\bar{x}_{geom\ s}$.

$$\bar{x}_{as}=\frac{4+6+7+10}{4}=6,75$$

$$\bar{x}_{geom\ s}=\sqrt[4]{4.6.7.10}=\sqrt[4]{1\ 680}=\sqrt{\sqrt{1\ 680}}=6,4.$$

Deci $\bar{x}_{as}$ dă rezultat mai mare fiind mai sensibilă la valorile extreme; mai aproape de adevăr este însă $x_{geom\ s}$.

Exemplul 3:

Fie valorile progresiei geometrice: 4; 8; 16; 32; 64. Să se afle termenul intermediar dintre 4 și 16 (presupunînd că nu-l cunoaștem).

$$\bar{x}_{as}=\frac{4+16}{2}=\frac{20}{2}=10.$$

$$\bar{x}_{geom\ s}=\sqrt{4\cdot 16}=\sqrt{64}=8.$$

Deci termenul intermediar real este 8.

Exemplul 4:

Să se determine termenul intermediar dintre 4 și 64.

$$\bar{x}_{as}=\frac{4+64}{2}=\frac{68}{2}=34$$

$\bar{x}_{geom\ s}=\sqrt{4\cdot 64}=\sqrt{256}=16$, așa cum se află în șirul dat.

Scanned with OKEN Scanner

Exemplul 5:

La o analiză de apă s-au găsit:

10 E. coli în prima zi,

1 000 E. coli în ziua a treia.

Care este media? (valoarea din ziua a doua)

$$\bar{x}_{as} = \frac{10+1\,000}{2} = 505$$

$\bar{x}_{geom\ s} = \sqrt[2]{10 \cdot 1\,000} = \sqrt[2]{10\,000} = 100$ E. coli în ziua a doua.

Exemplul 6:

La o analiză de apă s-au găsit:

100 E. coli în prima zi,

1 000 E. coli în ziua a treia.

Care este media? (valoarea din ziua a doua)

$$\bar{x}_{as} = \frac{100+1\,000}{2} = 1\,100:2 = 550$$

$$\bar{x}_{geom\ s} = \sqrt[2]{100 \cdot 1\,000} = \sqrt[2]{100\,000} = 316.$$

Deci prin $\bar{x}_{as}$ nu putem face deosebire între situațiile din cele două exemple (valoarea 505 fiind foarte apropiată de 550), ceea ce înseamnă că $\bar{x}_{as}$ nu este potrivită în astfel de situații. Deci $\bar{x}_{geom}$ se utilizează la calcularea ritmului mediu a fenomenelor ce evoluează, creșterea fiind regulată. Deci se utilizează cînd valorile se succed în ritm geometric (deci înmulțindu-se) nu în șir aritmetic (deci adăugîndu-se).

Exemplul 7:

Într-o localitate, în anul 1860 au fost 1 000 locuitori
„ 1880 „ 1 250 „
„ 1900 „ 1 562 „

Presupunînd că nu avem datele din 1880, să aflăm numărul locuitorilor.

$x_{geom\ s} = \sqrt{1\,000 \cdot 1\,562} = 1\,249$, adică atîția cîți au fost în 1880.

Exemplul 8:

Revenim la exemplul 5 de la $\bar{x}_{as}$: spitalul A=10 bolnavi/pat; spitalul B=13,3 bolnavi/pat; spitalul C=4 bolnavi/pat. Care este media?

$$\bar{x}_{geom\ s} = \sqrt[3]{10 \cdot 13{,}3 \cdot 4} = \sqrt[3]{532} = 8{,}1 \text{ bolnavi/pat}$$

Media geometrică ponderată

Se întrebuințează cînd valorile individuale nu au frecvențe egale.

Formula este: $\bar{x}_{geom\ p} = \sqrt[\Sigma f_1]{(x_1^{f_1}) \cdot (x_2^{f_2}) \cdots (x_n^{f_n})} = \sqrt[\Sigma f_1]{\pi x^f}$,

unde: π=„produsul dintre".

Exemplul 1:

La spitalul A=100 paturi cu 10 bolnavi/pat/an; spitalul B=150 paturi cu 13,3 bolnavi/pat/an; spitalul C=50 paturi cu 4 bolnavi/pat/an. Care este media? Problema se poate rezolva în două moduri:

a) $\bar{x}_{geom\ p} = (10^{100}) \cdot (13{,}3^{150}) \cdot (4^{50})$ și prin logaritmare avem:

$$= \frac{(100 \cdot \log 10) + (150 \cdot \log 13{,}3) + (50 \cdot \log 4)}{300} =$$

$$= \frac{(100 \cdot 1) + (150 \cdot 1{,}12385) + (50 \cdot 0{,}62)}{300} = \frac{100 + 168{,}5 + 30{,}1}{300} =$$

$$= \frac{298{,}6}{300} = 0{,}9955$$

Antilogaritmul mantisei 9 955=9 875, deci: 9,87 bolnavi/pat/an

b) $\bar{x}_{geom\ p} = 10^{\frac{100}{300}} \cdot 13{,}3^{\frac{150}{300}} \cdot 4^{\frac{50}{300}} = 10^{1/3} \cdot 13{,}3^{1/2} \cdot 4^{1/6} =$

$= 10^{0{,}33} \cdot 13{,}5^{0{,}5} \cdot 4^{0{,}16} =$

$= (0{,}33 \cdot \log 10) + (0{,}5 \cdot \log 13{,}3) + (0{,}16 \cdot \log 4) =$

$= 0{,}3333 + 0{,}56192 + 0{,}096329 = 0{,}99157$

Antilogaritmul mantisei 99157=9808, deci 9,8 bolnavi/pat/an.

Media geometrică procentuală

Se utilizează cînd valorile sînt date în procente. Formula este:

$$\bar{x}_{geom\ \%} = \sqrt[100]{(x_1^{p_1}) \cdot (x_2^{p_2}) \cdot \cdot \cdot (x_n^{p_n})} = \sqrt[100]{\pi(x)^p} \text{ unde:}$$

π = „produsul dintre".

Exemplul 1:

La spitalul A: 10 bolnavi au ocupat 33,3% din paturi; la spitalul B: 13,3 bolnavi au ocupat 50% din paturi; la spitalul C: 4 bolnavi au ocupat 16,7% din paturi. Care este media?

Problema se poate rezolva în două moduri:

a) $\bar{x}_{geom\ \%} = \sqrt[100]{(10^{33,3}) \cdot (13,3^{50}) \cdot (4^{16,7})}$ și logaritmînd avem:

$$= \frac{(33,3 \cdot \log 10) + (50 \cdot \log 13,3) + (16,7 \cdot \log 4)}{100} =$$

$$= \frac{(33,3 \cdot 1) + (50 \cdot 1,1238) + (16,7 \cdot 0,6020)}{100} = 0,9984$$

Antilogaritmul mantisei 9 984 = 9 965, deci: 9,96 bolnavi/pat/an.

b) $\bar{x}_{geom\ \%} = (10^{0,333}) \cdot (13,3^{0,5}) \cdot (4^{0,167})$ și logaritmînd avem:

$$= (0,333 \cdot \log 10) + (0,5 \cdot \log 13,3) + (0,167 \cdot \log 4) =$$

$$= (0,333 \cdot 1) + (0,5 \cdot 1,1238) + (0,167 \cdot 0,6020) =$$

$$= 0,333 + 0,5619 + 0,10053 = 0,9954$$

Antilogaritmul mantisei 9 954 = 9 895, deci: 9,8 bolnavi/pat/an.

Explicație: 0,333 + 0,5 + 0,167 = 1 = 100%

3.1.3. Media patratică

Media patratică simplă

Se întrebuințează cînd valorile prezintă creșteri din ce în ce mai mari, modificîndu-se aproximativ în progresie exponențială. Ea constituie modelul matematic pentru

„abaterea patratică". Media este sensibilă la valorile extreme, din care cauză este totdeauna mai mare decît celelalte medii. Are avantajul că se poate aplica și în cazul valorilor nule sau negative (care prin ridicare la patrat, devin pozitive). Se întrebuințează cînd dăm importanță valorilor mari. Se întrebuințează de ex. la calculul diametrului unui tub prin care trec gaze. Formula sa este următoarea:

$$\bar{x}_{patr\ s} = \sqrt{\frac{x_1^2 + x_2^2 + \cdots x_n^2}{n}} = \sqrt{\frac{\Sigma x_i^2}{n}}$$

Exemplul 1:

Fie valorile: 4; 10; 13,3. Care este media?

$$\bar{x}_{patr\ s} = \sqrt{\frac{4^2 + 10^2 + 13{,}3^2}{3}} \sqrt{97} = 9{,}8$$

Media patratică ponderată

Se întrebuințează cînd valorile nu au frecvențe egale. Formula este:

$$\bar{x}_{patr\ p} = \sqrt{\frac{(x_1^2 \cdot f_1) + (x_2^2 \cdot f_2) + \cdots (x_n^2 \cdot f_n)}{f_1 + f_2 + \cdots f_n}} = \sqrt{\frac{\Sigma\,(x^2 \cdot f)}{\Sigma\, f_i}}$$

Exemplul 1:

Fie valorile: 3; 3; 3; 4; 7; 7; 7. Care este media?

$$\bar{x}_{patr\ p} = \sqrt{\frac{(3^2 \cdot 4) + (4^2 \cdot 2) + (7^2 \cdot 3)}{4 + 3 + 2}} = \sqrt{\frac{215}{9}} = 4{,}88$$

Media patratică procentuală

Se întrebuințează cînd valorile sînt date în procente. Formula este următoarea:

$$\bar{x}_{patr\ \%} = \sqrt{\frac{(x_1^2 \cdot p_1) + (x_2^2 \cdot p_2) + \cdots (x_n^2 \cdot p_n)}{100}} = \sqrt{\frac{\Sigma\,(x^2 \cdot p)}{100}},$$

unde: p=frecvența procentelor.

Exemplul 1:

Fie valorile: 3 de 40%; 4 de 20%; 5 de 10%; 7 de 30%. Care este media?

Pentru demonstrație, întocmim tabelul:

valori x:	3	4	5	7
frecvențe %:	40	20	10	30=100%

$$\bar{x}_{patr\ \%}=\sqrt{\frac{(3^2\cdot 40)+(4^2\cdot 20)+(5^2\cdot 10)+(7^2\cdot 30)}{100}}=\sqrt{\frac{2\,400}{100}}=4{,}89$$

Exemplul 2:

La spitalul A 10 bolnavi au ocupat 33,3% din paturi

„ B 13,3 „ 50 „

„ C 4 „ 16,1 „

Care este media?

$$\bar{x}_{patr\ \%}=\sqrt{\frac{(10^2\cdot 33{,}3)+(13{,}3^2\cdot 50)+(4^2\cdot 16{,}7)}{100}}=\sqrt{\frac{12\,442}{100}}=$$

$$=11{,}1 \text{ bolnavi.}$$

3.1.4. Media armonică

Media armonică simplă

Exprimă caracterul sintetic al unor valori ce se află în raport invers. Se utilizează cînd frecvențele sînt egale. Se aplică atunci cînd nu avem frecvențele reale, ci produsele $x_i \cdot f_i$. De ex. la calculul vîscozității unor amestecuri de gaze. Rezultatul tinde către valorile mici. Rezultatul este mai mic decît la celelalte medii. Media armonică a 2 valori m_1 și m_2 se poate scrie și:

$\bar{x}_{arm\ s}=\frac{2(m_1+m_2)}{m_1+m_2}$. Formula este următoarea:

$$\bar{x}_{arm\ s}=\frac{n}{\frac{1}{x_1}+\frac{1}{x_2}+\ldots\frac{1}{x_n}}=\frac{n}{\Sigma\left(\frac{1}{x_i}\right)}$$

Exemplul 1:

Fie valorile: 1; 5; 10. Care este valoarea medie?

$$\bar{x}_{arm\ s} = \frac{3}{\frac{1}{1} + \frac{1}{5} + \frac{1}{10}} = \frac{3}{1,3} = 2,3$$

$$\bar{x}_{as} = (1+5+10) : 3 = 5,3$$

Deci vedem că $\bar{x}_{arm\ s}$ este mai mică decît $\bar{x}_{as}$.

Observaţie: în unele cărţi de statistică, este notată această medie $\bar{x}_{h\ s}$ (de la „harmonică").

Media armonică ponderată

Se întrebuinţează cînd ponderea nu se poate face prin $(x_i \cdot f_i)$, ci prin înmulţirea inverselor acestor valori. Se utilizează cînd frecvenţele nu sînt egale. Formula este următoarea:

$$\bar{x}_{arm\ p} = \frac{f_1 + f_2 + \cdot \cdot f_n}{\frac{f_1}{x_1} + \frac{f_2}{x_2} + \frac{f_n}{x_n}} = \frac{\Sigma f_i}{\Sigma \left(\frac{f_i}{x_i}\right)}$$

Exemplul 1:

În localitatea A cu 4 000 locuitori a fost 8% mortalitate

„	B	4 500	„	9%	„
„	C	7 000	„	10%	„

Care este media pentru cele trei localităţi?

$$\bar{x}_{as} = (8+9+10) : 3 = 9$$

$$x_{arm\ p} = \frac{4\,000 + 4\,500 + 7\,000}{\left(\frac{1}{8} \cdot 4\,000\right) + \left(\frac{1}{9} \cdot 4\,500\right) + \left(\frac{1}{10} \cdot 7\,000\right)} = \frac{15\,500}{17\,000} = 9,12\%$$

Media armonică procentuală

Se utilizează cînd datele sînt în procente. Formula este următoarea:

$$\bar{x}_{arm\ \%} = \frac{100}{\frac{p_1}{x_1} + \frac{p_2}{x_2} + \cdot \cdot \frac{p_n}{x_n}}, \text{ unde:}$$

p = procente

Scanned with OKEN Scanner

Exemplul 1:

Spitalul A cu 10 bolnavi a ocupat 33,3% paturi; spitalul B cu 13,3 bolnavi, 50% paturi, spitalul C cu 4 bolnavi, 16,7% paturi. Care este media?

$$\bar{x}_{arm\ \%} = \frac{100}{\frac{33,3}{10} + \frac{50}{13,3} + \frac{16,7}{4}} = \frac{100}{11,26} = 8,88 \text{ media.}$$

3.1.5. Media cronologică

Media cronologică simplă

Se utilizează cînd intervalele de timp dintre momentele seriei sînt egale. Formula este următoarea:

$$\bar{x}_{cron\ s} = \frac{\frac{x_1}{2} + x_2 + \cdots \frac{x_n}{2}}{n-1} = \frac{\frac{x_1 + x_n}{2} + x_i}{n-1}$$

Exemplul 1:

Într-o localitate au fost:

100	gravide	la începutul	trim. I
110	„	sfîrșitul	trim. I
123	„	„	II
132	„	„	III
80	„	„	IV.

Care este media gravidelor pe trimestre?

$$\bar{x}_{cron\ s} = \frac{\frac{100}{2} + 110 + 123 + 132 + \frac{80}{2}}{5-1} = \frac{455}{4} = 113,75$$

Exemplul 2:

Într-o localitate în anul 1860 au fost 1 000 locuitori

1880	„	1 250	„
1900	„	1 600	„
1920	„	1 950	„
1940	„	2 300	„

Scanned with OKEN Scanner

Care a fost numărul populației la mijlocul intervalului (anul 1900) presupunînd că nu avem cifra respectivă? Se consideră că nu au fost imigrări, emigrări, boli epidemice.

$$\bar{x}_{cron\ s} = \frac{\frac{1\,000}{2} + 1\,250 + 1\,950 + \frac{2\,300}{2}}{4-1} \quad \frac{4\,850}{3} = 1\,616$$

Media cronologică ponderată

Se utilizează cînd intervalele de timp dintre valorile șirului nu sînt egale. Formula este următoarea:

$$\bar{x}_{cron\ p} = \frac{\left(\frac{x_1+x_2}{2} \cdot f_1\right) + \left(\frac{x_2+x_3}{2} \cdot f_2\right) + \cdot\cdot \left(\frac{x_{n-1}+x_n}{2} \cdot f_n\right)}{f_1+f_2+\cdot\cdot f_n} = \frac{\left(\frac{x_1+x_{1n}}{2}\right) \cdot f_n}{\Sigma f_1}$$

Exemplul 3:

Într-un spital au fost la 1 iulie 107 bolnavi

1 nov. 108 „

31 dec. 100 „

Care a fost media zilnică de bolnavi?

$\bar{x}_{as} = \frac{107+108+100}{3} = 105$ bolnavi. Dar această medie nu este reală, deci trebuie să calculăm $\bar{x}_{cron\ p}$. Întocmim următorul tabel:

$$\left.\begin{array}{r} \text{iulie} = 31 \text{ zile} \\ \text{aug.} = 31 \text{ zile} \\ \text{sept.} = 30 \text{ zile} \\ \text{oct.} = 31 \text{ zile} \end{array}\right\} = 123 \text{ zile}$$

$$\left.\begin{array}{r} \text{nov.} = 30 \text{ zile} \\ \text{dec.} = 31 \text{ zile} \end{array}\right\} = 61 \text{ zile}$$

$$\bar{x}_{cron\ p}=\frac{\left(107\cdot\frac{123}{2}\right)+\left(108\cdot\frac{123+61}{2}\right)+\left(100\cdot\frac{61}{2}\right)}{\frac{123}{2}+\frac{123+61}{2}+\frac{61}{2}}=$$

$$=\frac{(107\cdot 61{,}5)+(108\cdot 92)+(100\cdot 30{,}5)}{61{,}5+92+30{,}5}=\frac{19\,566}{184}=106{,}3.$$

Diferența dintre $\bar{x}_{as}$ și $\bar{x}_{cron\ p}$ este dată de ponderea zilelor, care au fost calculate atît la numărător cît și la numitor.

3.1.6. Media glisantă

Numită și „medie mobilă", se utilizează în cazul în care șirul valorilor prezintă fluctuații mari, bruște și e greu de apreciat tendința („trend"-ul în termenii economiei comerciale), caracteristică a valorilor. Se presupune că media glisantă corespunde mijlocului intervalului sintetic. Calculul se poate face mediind 3 sau 5 valori alăturate. Se utilizează mai ales la construcția graficelor cu variații mari ale valorilor individuale. Formula sa este:

$$\bar{x}_{glis\ 3}=\frac{(x_i-1)+(x_i)+(x_i+1)}{3}=\frac{1}{3}\sum_{K=-1}^{K=i+1} x_i+k$$

$$\bar{x}_{glis\ 5}=\frac{(x_i-2)+(x_i-1)+(x_i)+(x_i+1)+(x_i+2)}{5}=\frac{1}{5}\sum_{K=-2}^{K=+2} x_i+k$$

Exemplul 1:

Fie valorile: 4; 10; 13,3.
Să se determine media glisantă 3.

$$\bar{x}_{glis\ 3}=\frac{4+10+13{,}3}{3}=9{,}1.$$

În cazul acesta, fiind numai 3 valori, $\bar{x}_{glis\ 3}=\bar{x}_{as}$.

Exemplul 2:

Fie valorile: 30; 20; 50; 40; 80; 20. Să se determine $\bar{x}_{glis\ 5}$.

$$x_{glis\ 5} = \frac{30+20+50+40+80}{5} = 44$$

$$= \frac{20+50+40+80+60}{5} = 50$$

$$= \frac{50+40+80+60+20}{5} = 50$$

Construind un grafic cu aceste valori (44; 50; 50) vedem că deşi variaţiile individuale x_i sînt mari ($\bar{x} \pm s = 46 \pm 24$), totuşi tendinţa „medie" a fenomenului este aproape staţionară.

3.1.7. Media patratelor

Fie următoarele viteze ale vîntului într-un interval de timp:

5 m/s; 7 m/s; 10 m/s. Care este media vitezelor?

$$\bar{x}_{as} = \frac{5+7+10}{3} = 7{,}3.$$

Dar, întrucît interesează acţiunea vîntului, care depinde de presiunea lui, aceasta este funcţie de patratul vitezei. Deci trebuie calculat astfel: $\bar{x}_{patrate} = \frac{5^2+7^2+10^2}{3} = \frac{178}{3} = 58\ m/s^2$, cifră care are utilitate în calculele referitoare la rezistenţa podurilor la vînt, la rezistenţa coşurilor de fabrică, la randamentul centralelor eoliene, la efectul asupra platformelor de foraj marin, la efectul asupra velelor ambarcaţiunilor etc., deoarece interesează efectul produs de vînt. Acest lucru se datorează faptului că $P_{kg/m^2} = v^2 \cdot 0{,}075$ şi redăm un astfel de tabel:

$V_{m/s}$	V^2	P_{kg/m^2}
1	1^2	$1 \cdot 0{,}075 = 0{,}075$
2	2^2	$4 \cdot 0{,}075 = 0{,}3$
3	3^2	$9 \cdot 0{,}075 = 0{,}67$
4	4^2	$16 \cdot 0{,}075 = 1{,}2$ etc.

Deci la triplarea vitezei vîntului, efectul creşte de 9 ori.

Scanned with OKEN Scanner

Un exemplu foarte practic este faptul că la autovehicule consumul de benzină crește cu patratul vitezei. Considerînd că o mașină consumă 5 l benzină la 50 km/h, avem următoarea situație:

KM/H	(KM/H)²	LITRI BENZINĂ/100 KM
50	2 500	5 litri
60	3 600	7,2
70	4 900	9,8
80	6 400	12,8
90	8 100	16,2
100	10 000	20

Deci la 100 km/h mașina consumă de 4 ori mai multă benzină decît la 50 km/h. Corelația nu e perfect liniară, depinzînd de mai mulți factori (tipul mașinii, drumul, încărcătura etc.), dar în general este valabilă relația. Calculul se face astfel:

2 500 (km/h)² ..... 5 litri
3 600 „ ..... x litri

$$x = \frac{3\,600 \cdot 5}{2\,600} = 7{,}2 \text{ litri.}$$

Consumul depinde și de „suprafața de înaintare" a vehiculului, de ex. un turism cu suprafața de înaintare de 1 m² la 60 km/h (30 km înaintare+30 km/h vînt contrar) va cheltui aprox. 3,4 H.P. (=aprox. 1,5 litri benzină), iar la 90 km/h va cheltui aprox. 14,6 H.P. (=aproximativ 6,5 litri benzină) numai pentru a învinge rezistența aerului, deoarece rezistența crește cu patratul vitezei. Deci amplasarea bagajelor pe capota mașinii are importanță și la consumul benzinei. Un alt exemplu este următorul: pentru a mări viteza unui avion de la 500 km/h la 1 000 km/h, trebuie mărită puterea motorului de 8 ori. De unde se vede că viteza costă scump!

Scanned with OKEN Scanner

3.1.8. Media progresivă

Într-un atelier sînt mașini care produc 10, 11, 12, 13, 14 piese pe zi. Care este media?

$\bar{x}_{as} = \frac{10+11+12+13+14}{5} = 12$ piese/zi. Dar, trebuie aplicată media progresivă a cărei formulă este următoarea:

$$\bar{x}_{progr} = \frac{\bar{x}_a + \bar{x}_s}{2},$$

unde:

$\bar{x}_a$ = media aritmetică

$\bar{x}_s$ = media termenilor calitativ superiori mediei aritmetice.

În cazul de față, avem: $\bar{x}_a = 12$

$$\bar{x}_s = \frac{13+14}{2} = 13{,}5$$

Deci: $\bar{x}_{progres} = \frac{12+13{,}5}{2} = 12{,}75.$

Logica este următoarea: dacă toate mașinile sînt de același tip, toate pot produce cel puțin 12,75 piese/zi, deoarece unele pot efectua chiar și 14 piese/zi. Deci cele ce produc mai puțin, trebuie reparate ca să ajungă la producția progresivă (12, 75 piese/zi).

3.1.9. Media logaritmică

Se întrebuințează pentru a stabili valoarea intermediară Q_i de la mijlocul intervalului de observație, cînd valoarea inițială Q_1 variază în decursul timpului după o lege exponențială (care este inversul funcției logaritmice), valoarea finală fiind Q_2. Formula sa este:

$$\bar{x}_{ln} = \frac{Q_1 - Q_2}{\ln Q_1 - \ln Q_2},$$

unde:

Q_1 = valoarea la începutul observației

Q_2 = valoarea la sfîrșitul observației

ln = logaritmul natural.

Scanned with OKEN Scanner

Exemplul 1:

O substanță radioactivă are $Q_1=20$ mCi/h, iar după 16 zile $Q_2=5$ mCi/h. Să se stabilească ce radioactivitate Q_i avea la mijlocul intervalului (la 8 zile).

$$\bar{x}_{ln}=\frac{Q_1-Q_2}{\ln\frac{Q_1}{Q_2}}=\frac{20-5}{\ln\frac{20}{5}}=\frac{15}{\ln 4}=\frac{15}{1{,}3863}=10{,}8 \text{ mCi/h.}$$

Exemplu 2:

Aceeași substanță radioactivă are $Q_1=10$ mCi/h, iar după 4 zile are $Q_2=7$ mCi/h. Care a fost Q_i la mijlocul intervalului (=2 zile)?

$$\bar{x}_{ln}=\frac{10-7}{\ln\frac{10}{7}}=\frac{3}{\ln 1{,}42857}=\frac{3}{0{,}35667}=8{,}41 \text{ mCi/h.}$$

Observație: În exemplul 1, valoarea exactă pentru $Q_i=10$, iar în exemplul 2 valoarea exactă pentru $Q_i=8{,}4$. De aici se vede că precizia determinării este mai mare cînd timpul de observație nu este prea lung și zecimalele sînt mai multe în calcul. Exemplele date se referă la Iod^{131} cu $T\frac{1}{2}=8{,}1$ zile.

3.1.10. Media mediilor

În unele cazuri, un șir lung de cifre se împarte în mai multe grupe pentru a se efectua calculele mai ușor. După ce se termină calculele parțiale, se face media generală, $\bar{\bar{x}}$.

Exemplul 1:

Fie următoarele valori: 10; 20; 30; 40; 50; 60; 70; 80; 90; 100. Să se determine $\bar{\bar{x}}$.

Acest șir fiind prea lung îl împărțim astfel:

	50+
	60
10+	70
20	80
30	90
40	100
100:4=25	450:6=75

Deci $\bar{\bar{x}}=(25+75):2=50$

Dar dacă șirul este împărțit astfel:

$$\begin{array}{l} 10+ \\ 20 \\ 30 \\ \hline 60:3=20 \end{array} \qquad \begin{array}{l} 40+ \\ 50 \\ 60 \\ 70 \\ 80 \\ 90 \\ 100 \\ \hline 490:7=70 \end{array}$$

Deci $\bar{\bar{x}}=(20+70):2=45$

Avem astfel două rezultate la aceeași populație și dacă împărțim șirul în alte porțiuni, obținem alte rezultate. Aceasta se datorează faptului că fiecare medie parțială intră în calcul fără ponderea ei.

Corect $\bar{\bar{x}}$ se stabilește astfel:

1. suma tuturor valorilor, împărțită la numărul valorilor;

2. împărțirea șirului în porțiuni egale, după care din mediile parțiale se calculează media $\bar{\bar{x}}$;

3. calcularea mediei ponderate, ex. $\bar{\bar{x}}=\frac{(20\cdot 3)+(70\cdot 7)}{3+7}=$ $=\frac{550}{10}=55.$

Deci corect este $\bar{\bar{x}}=55$.

3.1.11. Interpretarea rezultatelor și concluzii

Ca demonstrație, am rezolvat aceeași problemă utilizînd mai multe formule de calcul și redăm rezultatele obținute la problema Nr. 5 de la $\bar{x}_{as}$:

$\bar{x}_{as}$	$=9,1$	$\bar{x}_{glis\ 3}$	$=9,1$
$\bar{x}_{ap}$	$=10,6$	$\bar{x}_{patr\ s}$	$=9,8$
$\bar{x}_{a\ \%}$	$=10,6$	$\bar{x}_{patr\ p}$	$=9,8$
$\bar{x}_{geom\ s}$	$=8,1$	$\bar{x}_{patr\ \%}$	$=11,1$
$\bar{x}_{geom\ p}$	$=9,8$	$\bar{x}_{arm\ s}$	$=7$
$\bar{x}_{geom\ \%}$	$=9,9$	$\bar{x}_{arm\ p}$	$=8,8$
$\bar{x}_{cron}$	$=9,3$	$\bar{x}_{arm\ \%}$	$=8,8$

Scanned with OKEN Scanner

Deci rezultatele variază de la 7 la 11,1. Din acestea trebuie eliminate cîteva chiar de la început și anume:

$\bar{x}_{geom}$, deoarece valorile provin din locuri diferite, ce nu au nici o legătură între ele. Deci nefiind valori care să crească (prin înmulțire), metoda nu trebuie aplicată;

$\bar{x}_{patr}$ nu trebuie aplicată deoarece nu căutăm nici un indice. Fiind sensibilă la valorile mari, am obținut cifra cea mai mare;

$\bar{x}_{arm}$ nu trebuie aplicată deoarece valorile nu sînt în raport invers unele față de altele;

$\bar{x}_{cron}$ nu trebuie aplicată deoarece valorile nu variază în timp;

$\bar{x}_{glis}$ nu trebuie aplicată deoarece sînt numai 3 valori.

Deci, deși calculele sînt corecte d.p.d.v. matematic, rezultatul diferă, deoarece elementele ce au intrat în calcul au fost diferite; uneori au intrat în calcul numai bolnavii, alteori și paturile.

Rămîn deci în discuție $\bar{x}_{as}$ și $\bar{x}_{ap}$:

$\bar{x}_{as}$ reprezintă aici o medie a mediilor, care nu reflectă realitatea. Deci aceste medii parțiale nu trebuiau mediate astfel, fiecare medie parțială putînd servi, independent, pentru comparație;

$\bar{x}_{ap}$ reprezintă în cazul de față cel mai bine realitatea, căci ia în considerație fiecare factor care a influențat în vreun fel situația: număr de spitale, număr de paturi, număr de bolnavi.

Calculele respective se pot utiliza în diverse domenii de activitate, de ex. în loc de spitale și pacienți, se poate calcula producția în kg de pe 3 suprafețe agricole etc.

De aici rezultă că:

1. Media cea mai simplă de calculat este $\bar{x}_{as}$, dar rezultatul nu este totdeauna corect d.p.d.v. statistic;

2. Pentru fiecare problemă trebuie aleasă formula potrivită;

3. În probleme cu mai multe valori, trebuie utilizată media ponderată, care trebuie să fie egală cu media procentuală (eventualele diferențe fiind datorate zecimalelor, fracțiilor, procentelor care nu s-au calculat exact).

4. Raportul dintre aceste medii este:

$$\bar{x}_{patr} > \bar{x}_{a} > \bar{x}_{geom} > \bar{x}_{arm}.$$

3.2. Abaterea patratică

abatere standard, abatere medie patratică, abatere tip, abatere etalon, abatere sigma, écart quadratique moyen, standard deviation

S-a constatat că în cazul unei populații (obiecte, valori, experiențe, persoane etc.) mai mult sau mai puțin omogene, valorile individuale x_i sînt situate în jurul valorii $\bar{x}_{as}$ și anume, cu cît populația este mai mare, cu atît crește „șansa" de a fi mai multe valori x_i mai aproape de $\bar{x}_{as}$ („legea numerelor mari"). De aici rezultă că în cazul studierii unui fenomen, vom avea valori mai apropiate de medie, deci cu atît rezultatul este mai sigur, cu cît eșantionul este mai mare (în cazul că nu putem studia întreaga populație).

În concluzie, vom fi cu atît mai aproape de adevăr, cu atît mai mare va fi precizia determinării noastre, cu cît vor fi mai puține valori „răzlețe" față de medie. Și invers: cu cît „distanța" valorilor față de medie se abate mai mult, cu atît va fi mai inexact caracterizată populația (adică parametri populației). Calculul abaterii patratice derivă din media patratică, abaterea patratică ajutînd la evaluarea preciziei și siguranței că media este într-adevăr caracteristică pentru populația studiată, avînd și indicații asupra valorilor „răzlețe". Cu alte cuvinte, abaterea patratică ne arată gradul de „împrăștiere" a valorilor individuale x_i față de media $\bar{x}_{as}$.

Deoarece abaterea patratică se calculează atît cu numere pozitive, cît și cu numere negative, toate se ridică la patrat (cele negative devenind pozitive), după care se revine la cifrele reale prin extragerea rădăcinei patrate. Din această cauză se numește abatere „patratică". Această abatere se notează în calcule cu σ (sigma) cînd se ia în cercetare întreaga populație și cu „s" cînd se ia în cercetare numai un eșantion „s" (selectat) dintr-o populație.

Există și o abatere liniară $\bar{d}$ care nu se utilizează decît foarte rar, deoarece acordă aceeași importanță atît abaterilor mari, cît și abaterilor mici, astfel că nu se pot identifica factorii întîmplători care acționează. Formula

este următoarea: $\bar{d} = \frac{[x_1 - x_n]}{n}$, parantezele drepte însemnînd că valorile se iau în mărimi absolute. Din acest motiv, în calculele statistice se întrebuințează abaterea patratică, deoarece acordă o importanță mai mare abaterilor mari, fie că sînt cifre mici, fie că sînt cifre mari. Formula este următoarea:

$$s \text{ sau } \sigma = \sqrt{\frac{\Sigma (x_1 - \bar{x})^2}{n}} = \text{abatere patratică simplă}$$

$$s \text{ sau } \sigma = \sqrt{\frac{\Sigma (x_1 - \bar{x}) \cdot f}{f_1}} = \text{abatere patratică ponderată,}$$

unde:

x_i = valorile individuale

n = numărul cazurilor

f = frecvența cazurilor

f_i = frecvența tuturor cazurilor (numărul total al cazurilor).

Cînd sînt valori puține (de ex. sub 30 de cazuri) se întrebuințează o formulă „corectată" astfel:

$$s \text{ sau } \sigma = \sqrt{\frac{\Sigma (x_1 - \bar{x})^2}{n-1}}$$

Deci, din p.d.v. matematic, abaterea patratică este rădăcina patrată din media sumei deviațiilor de la medie. Cu alte cuvinte, este rădăcina patrată a împrăștierii valorilor x_i de la media $\bar{x}$.

Abaterea patratică este un indicator sensibil, care arată gradul de omogenitate al valorilor din colectivitate, în statistică (și în cercetările științifice), calcularea sa fiind obligatorie. Cine nu face acest lucru într-un studiu statistic înseamnă că are ceva de ascuns.

Calcularea abaterilor patratice se mai poate face și după următoarea formulă: $\sqrt{\frac{p(100-p)}{n}}$, unde:

p = procentajul.

Această formulă se va utiliza la testul t.

Scanned with OKEN Scanner

Exemplul 1:

înălțimea a 3 copii este: 90 cm, 100 cm, 110 cm. Care este media și abaterea patratică? În acest scop întocmim tabelul:

n	x_1
1	90
2	100
3	110
	300 : 3=100

Deci $\bar{x}_{as}=100$

Apoi calculăm, extinzînd tabelul:

n	x_1	$(x_1-\bar{x})$	$(x_1-\bar{x})^2$
1	90	90—100=—10	100
2	100	100—100=0	0
3	110	110—100=10	100
			200

Acum putem calcula abaterea patratică după formula:

$$s=\sqrt{\frac{\Sigma(x-\bar{x})^2}{n-1}}=\sqrt{\frac{200}{3-1}}=\sqrt{100}=10$$

Rezultatul se exprimă astfel: $\bar{x}\pm s=100\pm10$ cm.

Observații: am utilizat termenul s, deoarece cele 3 persoane sînt doar un eșantion dintr-o populație.

Interpretarea rezultatelor: media $\bar{x}$ se află la mijlocul abaterilor de ±10 cm, ceea ce se vede ușor aici, deoarece am luat numai 3 valori (din motiv didactic). În alte cazuri, cînd acest lucru nu se poate vedea, trebuie făcut calculul.

Scanned with OKEN Scanner

Exemplul 2:

Fie valorile: 70; 80; 90; 100; 110; 120; 130. Să se determine $\bar{x} \pm s$.

n	x_i	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$
1	70	70—100=—30	900
2	80	80—100=—20	400
3	90	90—100=—10	100
4	100	100—100=0	0
5	110	110—100=10	100
6	120	120—100=20	400
7	130	130—100=30	900
	700 : 7=100		2 800 : (7—1)=466,7

$\sqrt{466,7} = 21,6$

Rezultatul se exprimă: $\bar{x} \pm s = 100 \pm 21,6$.

Comparînd aceste rezultate observăm că:

1. mediile $\bar{x}_{as}$ sînt egale;
2. abaterile nu sînt egale. Aceasta înseamnă că omogenitatea populațiilor nu este egală, în exemplul 2 valorile sînt mult mai disparate: între cel mai mic și cel mai mare există o diferență de 60 unități.

Exemplul 3:

Fie valorile: 100; 100; 100; Să se determine $\bar{x} \pm s$.

n	x_i	$(x_i - \bar{x})$	$(x - \bar{x})^2$
1	100	100—100=0	0
2	100	100—100=0	0
3	100	100—100=0	0
	300 : 3=100		0 : 3—1=0

Rezultatul este $\bar{x} \pm s = 100 \pm 0$.

Deci nu este nici o abatere de la medie, deoarece toate valorile sînt egale.

Scanned with OKEN Scanner

Exemplul 4:

Fie valoarea: 100. Să se determine $\bar{x} \pm s$.

Într-un caz unic nu se poate determina nici o abatere patratică, deoarece nu avem cu cine face comparație.

Exemplul 5:

Fie următoarele valori ale temperaturii într-un interval de timp:

—10°C; 0°C; +10°C. Să se determine $\bar{x} \pm s$.

n	x_1	$(x_1-\bar{x})$	$(x-\bar{x})^2$
1	—10	—10—0=—10	100
2	0	0—0=0	0
3	+10	10—0=10	100
	0 : 3=0		200 : (3—1)=100
			$\sqrt{100}=10$

Rezultatul este $\bar{x} \pm s = 0 \pm 10$°C.

Exemplul 6:

Fie valorile: 10; 10; 10; 20; 20; 30. Să se determine $\bar{x} \pm s$.

După metoda clasică se procedează astfel:

n	x_1	$(x_1-\bar{x})$	$(x_1-\bar{x})^2$
1	10	10—16,6=6,6	43,5
2	10	10—16,6=6,6	43,5
3	10	10—16,6=6,6	43,5
4	20	20—16,6=3,4	11,5
5	20	20—16,6=3,4	11,5
6	30	30—16,6=13,4	179,5
6	100		335

$$x = \frac{100}{6} = 16{,}6$$

$$s = \sqrt{\frac{\Sigma (x-\bar{x})^2}{n-1}} = \sqrt{\frac{335}{6-1}} = \sqrt{67} = 8{,}1$$

Scanned with OKEN Scanner

Calcularea se mai poate face și după metoda, „calculator“, deoarece unele calculatoare au un astfel de program implementat prin fabricație. Se procedează astfel:

n	(x_i)	(x_i^2)
1	10	100
2	10	100
3	10	100
4	20	400
5	20	400
6	30	900
6	100	2 000

$$\bar{x}=100:6=16,6.$$

$$s=\sqrt{\frac{\Sigma(x_i^2)-\frac{\Sigma(x_i)^2}{n}}{n-1}}=\sqrt{\frac{2\,000-\frac{100^2}{6}}{6-1}}=\sqrt{\frac{2\,000-\frac{10\,000}{6}}{5}}=$$

$$=\sqrt{\frac{2\,000-1\,666}{5}}=\sqrt{\frac{334}{5}}=\sqrt{66,8}=8,1.$$

Deci rezultatul este la fel ca și după metoda clasică.

Exemplul 7:

Aceste valori pot fi „comprimate“ astfel: 3 valori de 10; 2 valori de 20; 1 valoare de 30.

În acest caz se calculează după metoda ponderată, astfel:

f	$\cdot x_i$	$= f\cdot x_i$	$(x_i-\bar{x})$	$(x_i-\bar{x})^2\cdot f=$
3	$\cdot 10$	$=30$	$10-16,6=6,6$	$43,5\cdot 3=130,6$
2	$\cdot 20$	$=40$	$20-16,6=3,4$	$11,5\cdot 2=23,1$
1	$\cdot 30$	$=30$	$30-16,6=13,4$	$179,5\cdot 1=179,5$
6		100		333,2

$$\bar{x} = \frac{(x_i \cdot f)}{f} = \frac{100}{6} = 16{,}6.$$

$$s = \sqrt{\frac{\Sigma (x_i - \bar{x})^2 \cdot f}{f-1}} = \sqrt{\frac{333{,}2}{6-1}} = \sqrt{66{,}6} = 8{,}1.$$

Rezultatul este: $\bar{x} \pm s = 16{,}6 \pm 8{,}1$.

Rezolvarea se poate face după metoda „calculator", astfel:

	$f \cdot x_i = f \cdot x_i$	$x_i^2 \cdot f$
3	$\cdot 10 = 30$	$100 \cdot 3 = 300$
2	$\cdot 20 = 40$	$400 \cdot 2 = 800$
1	$\cdot 30 = 30$	$900 \cdot 1 = 900$
6	100	2 000

$$\bar{x} = \frac{\Sigma (x \cdot f)}{f} = \frac{100}{6} = 16{,}6.$$

$$s = \sqrt{\frac{\Sigma (x_i^2 \cdot f) - \frac{\Sigma (x_i \cdot f)^2}{f}}{f-1}} = \sqrt{\frac{2\,000 - \frac{100^2}{6}}{6-1}} =$$

$$= \sqrt{\frac{2\,000 - \frac{10\,000}{6}}{5}} = \sqrt{\frac{2\,000 - 1\,666}{5}} = \sqrt{66{,}6} = 8{,}1.$$

Și în acest caz, rezultatele sînt egale.

3.2.1. Interpretarea rezultatelor și aplicații practice

Abaterea patratică servește drept „coeficient de precizie" statistică asupra valorilor individuale x_i, după care ne putem da seama ce încredere putem acorda rezultatului. Ea este un numitor comun, universal, pentru toate unitățile, de orice natură ar fi ele. Se exprimă în aceleași unități de măsură, ca și $\bar{x}$: dacă $\bar{x}$ reprezintă kg, abaterea

patratică reprezintă de asemenea kg etc. În toate cazurile, numărul valorilor (persoane, experiențe, măsurători etc.) are influență, observațiile arătînd că trebuie să avem peste 30 de valori pentru a avea încredere în rezultat. Datele, cifrele, valorile, se înregistrează în ordinea apariției lor (în ordine cronologică) pentru a vedea evoluția fenomenului în timp. Dacă nu interesează acest lucru, valorile se pot grupa în ordine crescîndă sau descrescîndă, cum este mai comod pentru calcul. Concordanța dintre determinări și precizia lor se poate vedea grafic astfel: coloanele sînt proporționale cu $\bar{x}$-urile, iar s-urile sînt reprezentate printr-o linie proporțională cu valoarea. Astfel se poate aprecia vizual diferența dintre $\bar{x}$-uri și s-uri (fig. 5).

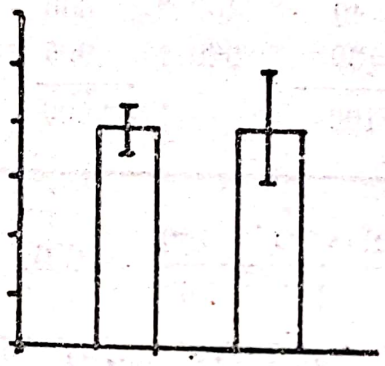


FIG. 5.

Ca aplicații practice redăm cîteva exemple:

1. Glicemia omului sănătos este 100 ± 10 mg la 100 ml sînge, aceasta înseamnă că valori de 90 . . . 110 mg se află la foarte mulți oameni sănătoși.

2. Metabolismul bazal este considerat normal dacă la repetate determinări se găsesc valori de $\pm 10\%$ față de media metabolismului bazal standard, determinat la un număr mare de oameni sănătoși. Aceste variații se datoresc anturajului, emoțiilor, ambianței termice, efortului fizic etc.

3. Timpul de erupție a dinților la copii este în funcție de vîrstă și de o serie de alți factori variabili: ereditate, boli, alimentație etc. (fig. 6).

Reprezentarea grafică a timpului mediu $\bar{x}$ și a abaterilor s ne fac să înțelegem că este inutilă medicația „stimulatoare", iar dacă nu este mugure dentar, orice medicație nu are rost. Din grafic rezultă că un copil de 20 de

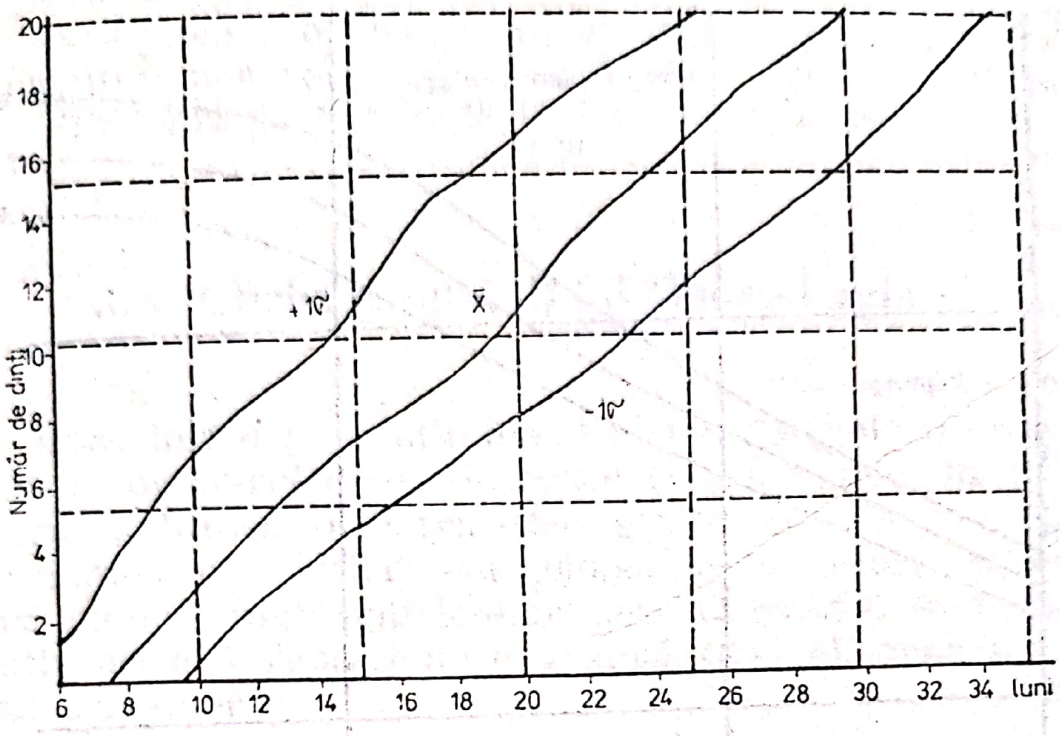


FIG. 6.

luni are în medie 10 dinți, dar nu e nimic grav dacă are 7—16 dinți (Fig. 6).

4. Dezvoltarea fizică a copiilor se supraveghează prin măsurători antropometrice, care se compară cu media altor măsurători pe un număr de copii de același sex, vîrstă, mediu etc., care constituie etalonul. Reprezentarea grafică (fig. 7) este construită pe baza unor măsurători în masă, după care s-a calculat și s-a trasat grafic $\bar{x} \pm s$. Pe un astfel de grafic se înscriu valorile obținute la măsurătorile anuale ale copiilor, obținîndu-se curbe-funcții pentru înălțime, greutate, perimetru toracic. Dacă aceste curbe se situează în limitele ± 1 sigma, înseamnă că dezvoltarea este armonioasă. Dacă vreuna din aceste curbe iese din „intervalul de siguranță" de ± 1 sigma trebuie luate măsuri de îndreptare: copilul este prea slab, prea mic etc. Din grafic se vede că un copil de 15 ani poate avea înălțimea între 155 și 173 cm (± 1 sigma). Un astfel de grafic este ușor de înțeles de elevi, care vor colabora activ cu medicul școlar în vederea „normalizării" valorilor lor. Deoarece aceste valori ($\bar{x}$ sau σ) se modifică în decursul timpului, pe grafic trebuie notat anul la care se referă etalonul.

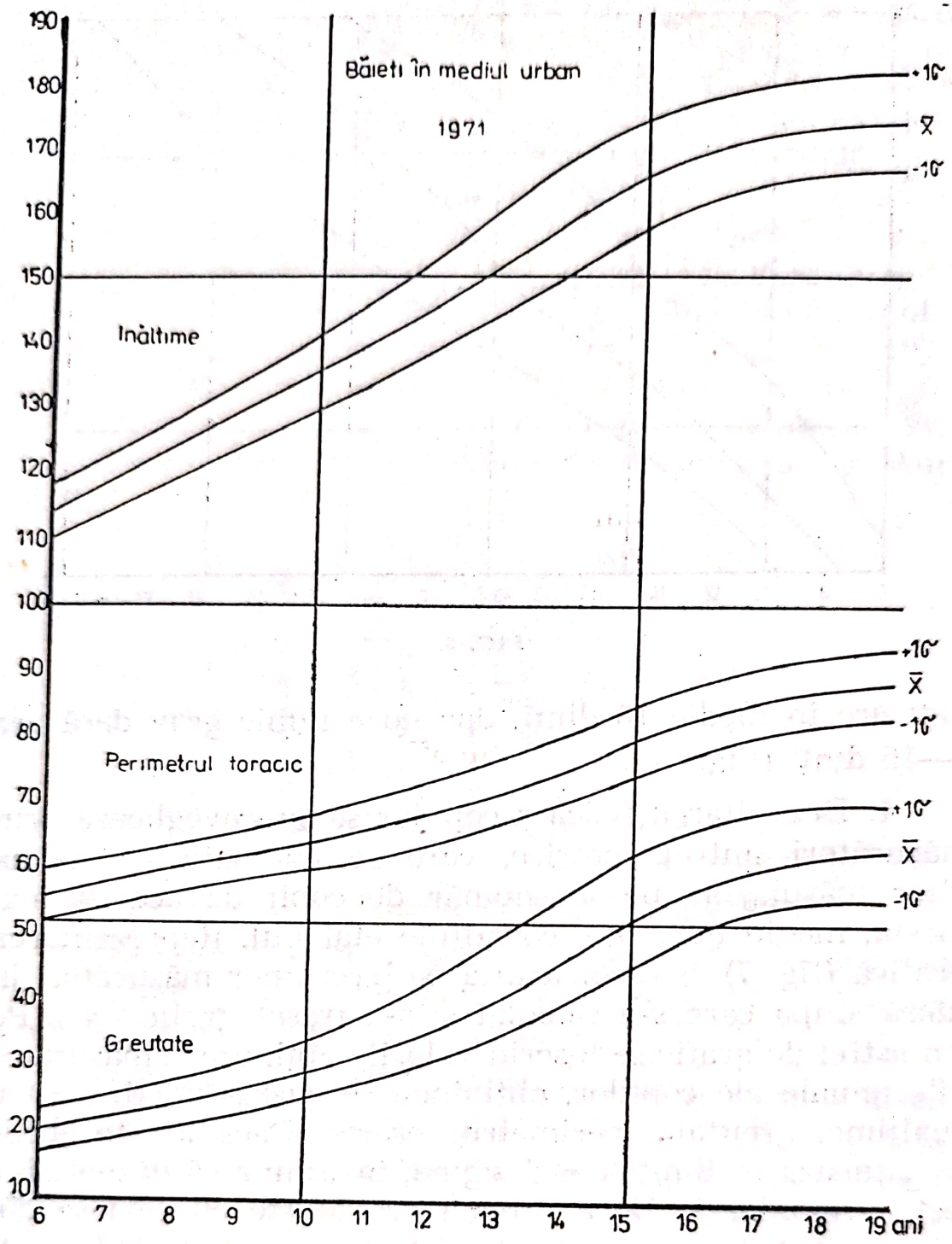


FIG. 7.

Comentarii:

— Calcularea lui σ sau s este o „garanție" a acurateții determinărilor; abaterea patratică se poate calcula numai dacă există 3 valori;

— Abaterea patratică este o valoare concluzivă numai dacă provine din cel puțin 30 valori;

— Abaterea patratică se calculează numai pentru o distribuție „normală“, „gaussiană“ sau „aproape gaussiană“. Problema va fi tratată la Curba lui Gauss.

3.3. Curba (clopotul) lui Gauss-Laplace

Dacă într-o populație avem elemente (valori) crescătoare și descrescătoare, simetrice și progresive, în jurul valorii $\bar{x}$, obținem o reprezentare grafică ce este cunoscută sub numele de „curba“ sau „clopotul“ lui Gauss-Laplace, după numele matematicienilor care au studiat mai amănunțit această reprezentare grafică (fig. 8), formula sa fiind următoarea:

$$y = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(x-\bar{x})^2}{2\sigma^2}}$$

În mod uzual, această curbă sau ogivă se reprezintă ca în fig. 9 și este denumită prescurtat „curba lui Gauss“. Astfel de situații se întîlnesc în toate cazurile de mulțimi (populații) cu distribuție „normală“ sau „întîmplătoare“ (deci unde nu s-a făcut vreo selecție), numită din această cauză „gaussiană“. Pentru explicarea problemei, dăm cîteva exemple:

— majoritatea formelor de boală sînt de gravitate medie; totuși există unele cazuri ușoare, iar altele grave;

— într-o sală de clasă, majoritatea elevilor (care au aproximativ aceeași vîrstă) au cam aceeași înălțime și sînt așezați în băncile din mijloc; totuși cîțiva elevi mai scunzi sînt așezați în băncile din față, iar cîțiva mai înalți sînt așezați în ultimele bănci;

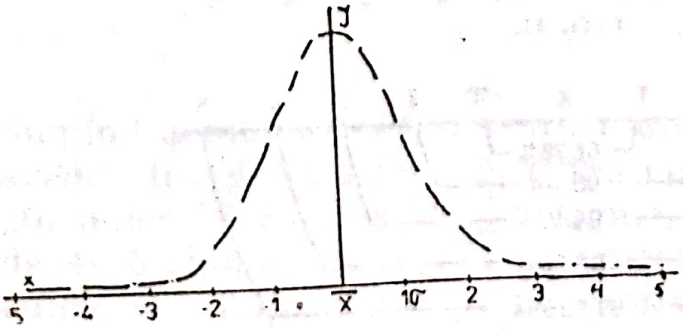


FIG. 8.

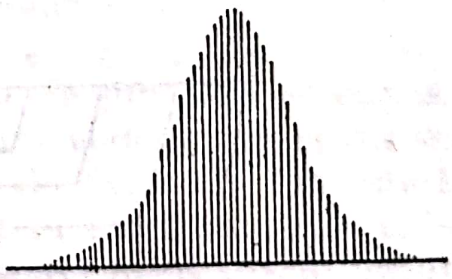


FIG. 9.

— pe sticlele de bere scrie „500 ml±2%“, deci majoritatea sticlelor conțin 500 ml, dar cîteva +1 . . . 2%, alte cîteva —1 . . . 2%.

Acest lucru se datorează faptului că obiectele, ființele, experiențele etc. nu sînt absolut egale între ele, deci există variații în plus și în minus față de medie (un aforism al lui Bernard Shaw spune că „în toate turmele sînt și oi negre“). Deci curba lui Gauss este reprezentarea grafică a abaterilor valorilor x_i față de x. Un studiu asupra acestei repartiții „normale“ apare pentru prima dată într-o lucrare a lui Abraham de Moivre (1667—1754), scrisă în limba latină. În anul 1794 ea este reluată de matematicianul și astronomul francez Pierre Simon de Laplace (1749—1827), dar importanța sa este arătată mai ales în lucrările lui Karl Friederich Gauss (1777—1855), matematician și astronom german, căutînd legea la care se supun

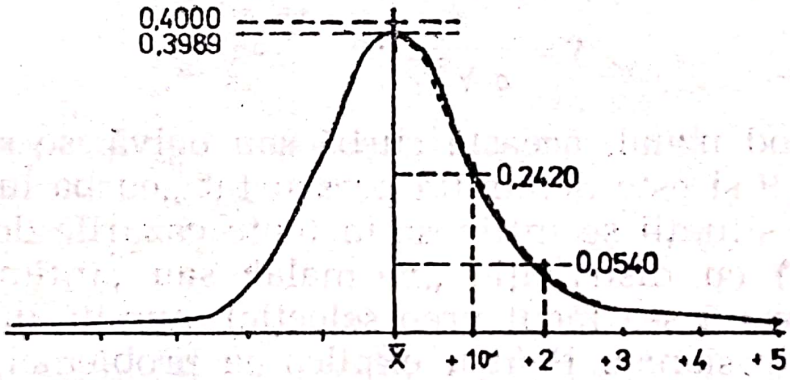


FIG. 10.

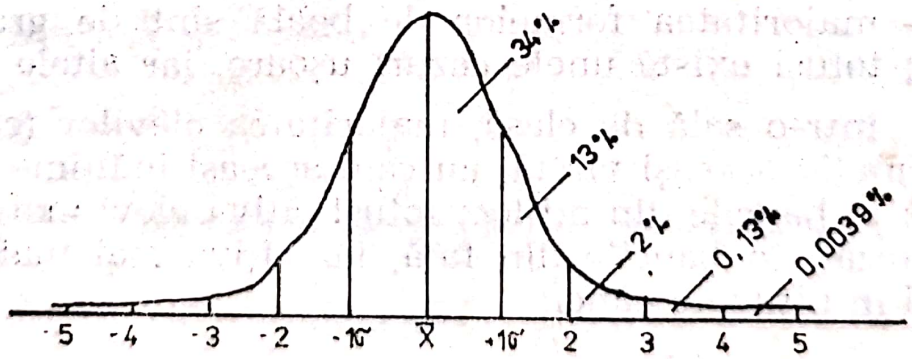


FIG. 11.

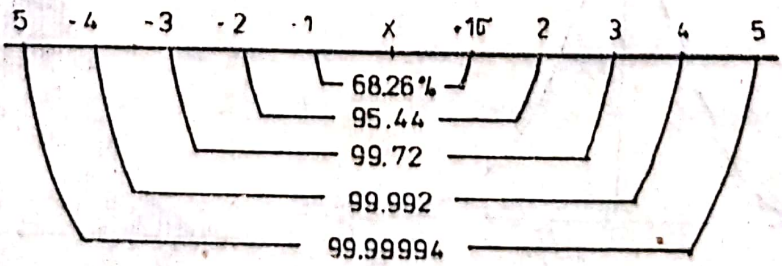


FIG. 12.

erorile de observație. Această curbă, numită și „curba erorilor mari“ reprezintă repartiția „normală“, „întîmplătoare“ sau „gaussiană“, spre deosebire de o repartiție inegală (repartiția „cazurilor rare“), care corespunde la „legea numerelor mici“, ce se rezolvă cu o formulă dată de Denis Simeon Poisson, stabilită cu aproape un secol în urmă, avînd atunci aplicativitate la urmăririle judiciare, anchetele criminale etc.

O curbă tipică se poate construi după coordonatele date în tabelul anexă, mai simplu decît după formula matematică (fig. 10). Aici se vede că:

înălțimea maximă $=0{,}3989$ (deoarece $\frac{1}{\sqrt{2\pi}}=0{,}3989$)

la 1 sigma $=0{,}2420$

la 2 sigma $=0{,}0540$, etc.

Dar curba lui Gauss ne interesează pentru suprafața circumscrisă între curbă și linia de bază, suprafață ce corespunde la elementele, valorile, numărul populației etc., luate în studiu. Din acest motiv suprafața de sub curbă este împărțită în mai multe „felii“ sau „porțiuni“ corespunzătoare abaterilor sigma. În fig. 11 sînt reprezentate aceste suprafețe, unde se vede că:

între $\bar{x}$ și $+1$ sigma $=34{,}3\%$ din totalul suprafeței
între $+1$ și $+2$ sigma $=13{,}5\%$ din totalul suprafeței
etc.

Totalizînd aceste suprafețe, avem:

$$\begin{array}{r} 34{,}13 \quad + \\ 13{,}59 \\ 2{,}14 \\ 0{,}136 \\ 0{,}00397 \\ \hline 49{,}99997 \end{array}$$

Sumînd aceste suprafețe pentru ambele jumătăți de curbă, avem: situația din fig. 12. După cum vedem, curba nu se „închide“ nici chiar la ± 5 sigme, dar valorile mai mari de 4—5 sigme nu se iau de obicei în considerare în calculele obișnuite, deoarece nu reprezintă decît foarte puțin din totalul valorilor, cu alte cuvinte, aceste elemente rare

Scanned with OKEN Scanner

nu influențează „mersul general“ al evenimentelor. Proprietățile curbei lui Gauss sînt următoarele:

1. Cînd populația este „normală“, forma curbei este tipică (fig. 13 a). Cînd populația a fost selecționată în vreun fel, curba este mai ascuțită (fig. 13 b), iar cînd sînt abateri mari față de $\bar{x}$, curba este mai aplatizată (fig. 13 c).

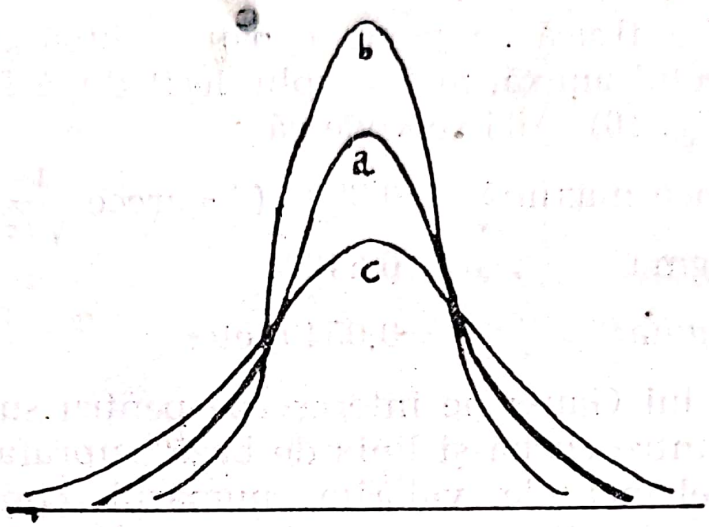


FIG. 13 a, b, c.

2. Unii autori consideră că o populație „normală“ (pentru care să se poată construi o curbă Gauss) trebuie să aibă cel puțin 30 elemente (alții pretind 40 sau chiar 60). În orice caz, cu cît avem mai multe elemente, cu atît avem șansa ca majoritatea lor să aibă proprietăți asemănătoare și să fie grupate mai aproape de $\bar{x}$. Cînd populația are elemente puține, în calcule se întrebuințează testul t.

3. Abaterea sigma ne arată:

— gradul de omogenitate (sau, din contră, de „împrăștiere“) a valorilor x_i față de $\bar{x}$;

— proporția bine definită a abaterilor față de $\bar{x}$, deoarece ± 1 sigma înseamnă 68,3% din populație etc. În acest caz, doar 100—68=32% din rezultatele noastre diferă mult de $\bar{x}$ etc.

Exemplul 1:

După măsurători și calcule s-a stabilit că într-o clasă înălțimea copiilor este 100 ± 5 cm. Măsurînd un copil din aceeași școală, găsim că are 104 cm. În acest caz, consi-

derăm că acest copil poate face parte din această clasă (teoretic). Măsurînd un alt copil, găsim că are 120 cm înălțime. Aceasta înseamnă că are 20 : 5 = +4 sigme în plus față de $\bar{x}$ a clasei. Din această cauză putem considera că acest copil nu face parte din clasa respectivă, deoarece:

în domeniul ± 1 sigma sînt cuprinși 68% din copii, deci 100—68=32% au alte înălțimi;

în domeniul ± 2 sigma sînt cuprinși 95,5% din copii, deci 100—95,5=4,5% au alte înălțimi;

în domeniul ± 3 sigma sînt cuprinși 99,7% din copii, deci 100—99,7=0,3%=3‰ din copii au alte înălțimi;

în domeniul ± 4 sigma sînt cuprinși 100—99,99= =0,01%=1 la 10 000 copii de aceeași vîrstă, sex etc., care au alte înălțimi pentru aceeași clasă.

Exemplul 2:

La școala A s-au obținut următoarele valori la măsurătorile făcute:

7 elevi au avut înălțimea 92 cm,

10 elevi au avut înălțimea 94 cm,

11 elevi au avut înălțimea 96 cm,

14 elevi au avut înălțimea 98 cm,

16 elevi au avut înălțimea 100 cm,

14 elevi au avut înălțimea 102 cm,

11 elevi au avut înălțimea 104 cm,

10 elevi au avut înălțimea 106 cm,

7 elevi au avut înălțimea 108 cm.

Să se calculeze $\bar{x} \pm s$ și să se reprezinte grafic rezultatul.

Scanned with OKEN Scanner

Calculul este următorul:

$f \cdot x_1$	=	fx_1	$(x_1-\bar{x})^2$	=	
7 · 92	=	644	92—100 = $8^2 \cdot 7$	=	448
10 · 94	=	940	94—100 = $6^2 \cdot 10$	=	360
11 · 96	=	1 056	96—100 = $4^2 \cdot 11$	=	176
14 · 98	=	1 372	98—100 = $2^2 \cdot 14$	=	56
16 · 100	=	1 600	100—100 = 0 · 16	=	0
14 · 102	=	1 426	102—100 = $2^2 \cdot 14$	=	56
11 · 104	=	1 144	104—100 = $4^2 \cdot 11$	=	176
10 · 106	=	1 060	106—100 = $6^2 \cdot 10$	=	360
7 · 108	=	756	108—100 = $8^2 \cdot 7$	=	448
100		10 000			2 080

$$\bar{x} = \frac{\Sigma (x \cdot f)}{f} = \frac{10\,000}{100} = 100 \text{ cm}$$

$$s = \sqrt{\frac{\Sigma (x_1 - \bar{x})^2}{f}} = \sqrt{\frac{2\,080}{100}} = \sqrt{20{,}8} = 4{,}56$$

Rezultatul este: $\bar{x} \pm s = 100 \pm 4{,}56$ cm

După metoda „calculator“ se procedează astfel:

$f \cdot x_1$	=	$f \cdot x_1$	x^2	$(x_i^2) \cdot f$
7 · 92	=	644	8 464	59 248
10 · 94	=	940	8 836	88 360
11 · 96	=	1 056	9 216	101 376
14 · 98	=	1 372	9 604	134 456
16 · 100	=	1 600	10 000	160 000
14 · 102	=	1 428	10 404	145 656
11 · 104	=	1 144	10 816	118 976
10 · 106	=	1 060	11 236	112 360
7 · 108	=	756	11 664	81 648
100		10 000		1 002 080

$$\bar{x} = \frac{x_1 \cdot f}{f} = \frac{10\,000}{100} = 100$$

$$s=\sqrt{\frac{\Sigma(x_i^2)\cdot f-\frac{(x_i\cdot f)^2}{f}}{f}}=\sqrt{\frac{1\,002\,080-\frac{(10\,000)^2}{100}}{100}}=$$

$$=\sqrt{\frac{1\,002\,080-\frac{10\,000\,000}{100}}{100}}=\sqrt{\frac{1\,002\,080-100\,000}{100}}=$$

$$=\sqrt{\frac{080}{100}}=\sqrt{20{,}8}=4{,}56$$

Rezultatul este același: $100\pm4{,}56$.

Stabilirea coordonatelor se face astfel (pentru grafic):

— pe ordonata x se iau 16 dimensiuni (=16 elevi);

— pe abscisa y se iau 16 dimensiuni (108 cm—92 cm=16 valori) (Fig. 14), coordonate pe care se transpun valorile individuale x_i și în funcție de frecvența f se trasează curba (fig. 15). Dacă la școala B obținem altă curbă, mai ascuțită, înseamnă că s-a făcut o selecție, de ex. pentru a alcătui o echipă sportivă.

Curba lui Gauss are aplicații și în pedagogie: majoritatea elevilor au note medii, cîțiva au note foarte bune și . . . cîțiva au note mici. O experiență (fig. 16) arată că profesorul „sever" nu admite nici o greșeală, deci notarea lucrărilor scrise este destul de subiectivă, profesorul sever acordînd în general note mai mici (fig. 17). Din această cauză, tendința în învățămîntul programat este ca notele să fie acordate de calculator, care este absolut obiectiv. Dezvoltarea anticorpilor în organism se face, de asemenea, după o curbă Gauss, de ex. în urma unei vaccinări. Deplasarea curbei în funcție de metabolismul bazal denotă afecțiuni medicale diferite (fig. 18). Unele aparate de laborator înregistrează rezultatele analizelor sub forma unei curbe Gauss. Dar uneori, pentru a uniformiza rezultatele, un buton de „atenuare" aplatizează curba, sau, din contră, pentru a stabili nu punct precis (de ex. la titrare), curba trebuie transformată în pante bruște, acționînd asupra altui buton.

Scanned with OKEN Scanner

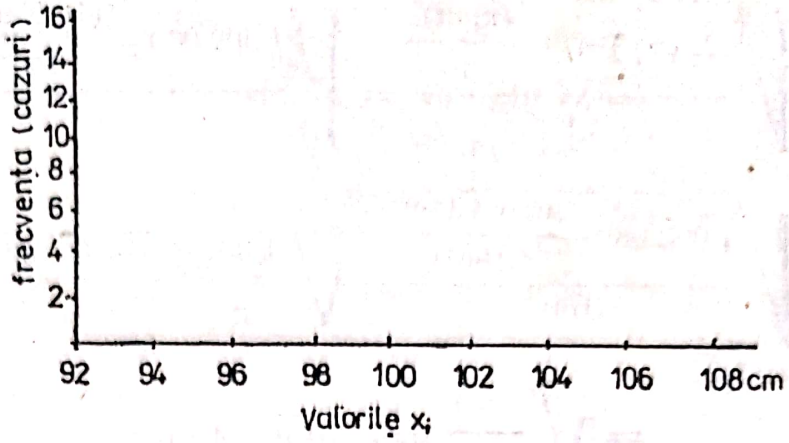


FIG. 14.

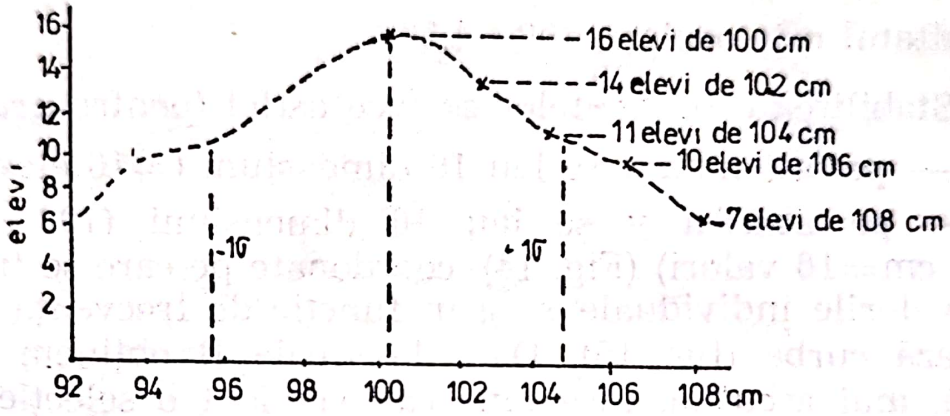


FIG. 15.

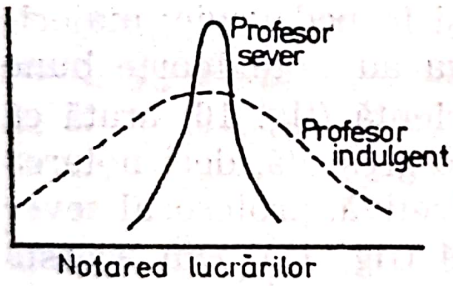


FIG. 16.

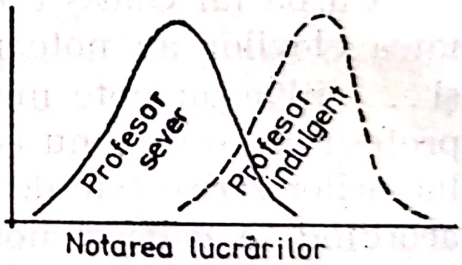


FIG. 17.

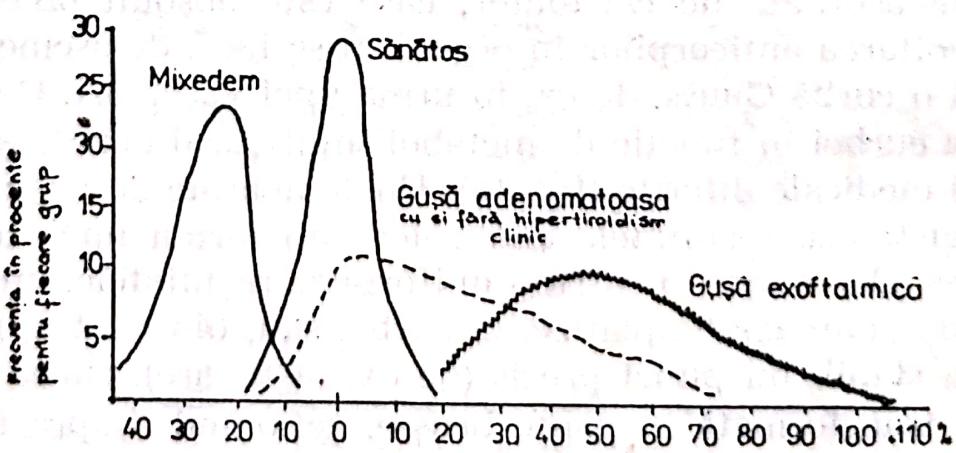


FIG. 18.

Concluzii:

1. Repartiția „normală“ este simetrică. Micile asimetrii nu schimbă proprietățile de repartiție „normală“. Erorile pozitive și negative sînt la fel de frecvente.

2. Repartiția „normală“ are un singur maxim (pe grafic). Cînd există două maxime, este probabilă amestecarea a două eșantioane diferite.

3. Cînd abaterea s este mică, valorile aleatorii mari sînt mai puțin frecvente decît valorile aleatorii mici.

4. Valorile x_i ce depășesc ± 3 s, de obicei, nu trebuie incluse în calcule: ele trebuie tratate ca atare, fiind cazuri „rare“.

5. Abaterile patratice sînt simetrice la distribuțiile gaussiene (ex. 10 ± 2), spre deosebire de abaterile altor distribuții (cînd poate fi de ex. 10^{+4}_{-2}).

3.4. Variabilitatea mediilor

Cînd dispunem de o populație întreagă, îi putem determina parametri $\bar{\bar{x}} \pm \sigma$. Dacă din această populație nu dispunem decît de unul sau mai multe eșantioane, le putem determina parametri $\bar{x}_A \pm s_A$; $\bar{x}_B \pm s_B$; $\bar{x}_C \pm s_C$ etc. Se va observa că valorile eșantioanelor variază, ele nu sînt egale între ele și nici cu $\bar{\bar{x}} \pm \sigma$, decît în cazul întîmplător cînd am avut în eșantion o parte din populație, care să aibă parametri egali cu ai întregii populații (fig. 19).

Reprezentînd grafic (fig. 20) constatăm că mediile $\bar{x}$ ale eșantioanelor se repartizează grafic (și variază ca va-

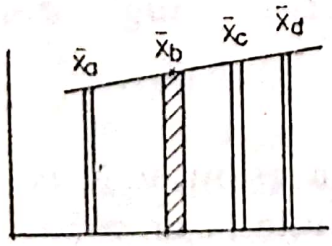


FIG. 19.

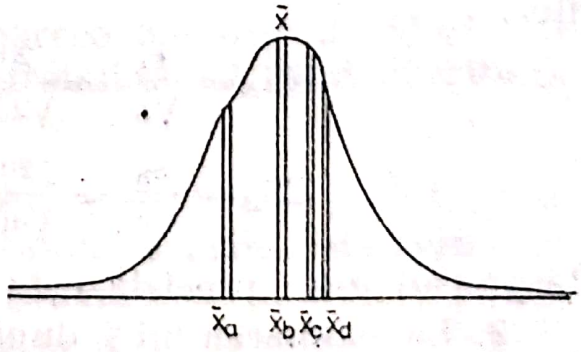


FIG. 20.

loare) în jurul mediei generale $\bar{\bar{x}}$, în cazul unei distribuții „normale“ pe curba lui Gauss.

3.4.1. Eroarea Standard ES sau $\sigma_{\bar{x}}$

Probabilitatea ca rezultatul $\bar{x} \pm s$ să fie diferit mai mult sau mai puțin de $\bar{\bar{x}} \pm \sigma$ depinde de numărul de elemente n ce am avut în eșantion. Deci cu cît avem mai multe elemente, cu atît ne apropiem de numărul total al populației. Corelația dintre abaterea s și numărul elementelor n din eșantion este dată de formula: $ES = \frac{s}{\sqrt{n}}$, unde: s = abaterea eșantionului și n = numărul elementelor din eșantion. De fapt, probabilitatea preciziei depinde de $\sqrt{n}$ și cînd numărul elementelor n este mic (ex. sub 30 de cazuri) formula utilizată este: $ES = \frac{s}{\sqrt{n-1}}$. Calculul lui ES se mai poate face și după formula:

$$ES = \sqrt{\frac{(x-\bar{x})^2}{n(n-1)}}$$

ES are mai multe întrebuințări și anume:

1. La stabilirea raportului dintre precizia unor determinări.

Exemplul 1:

Doi cercetători au întrebuințat metode diferite de lucru și au obținut:

$\bar{x}_A \pm s_A = 100 \pm 20$ din $n = 200$ elemente;
$\bar{x}_B \pm s_B = 90 \pm 20$ din $n = 400$ elemente.

Să se stabilească raportul dintre precizia determinărilor.

$$ES_A = \frac{s_A}{\sqrt{n_A}} = \frac{20}{\sqrt{200}} = \frac{20}{14,1} = 1,41$$

$$ES_B = \frac{s_B}{\sqrt{n_B}} = \frac{20}{\sqrt{400}} = \frac{20}{20} = 1$$

Răspunsul este: precizia este mai mare la eșantionul B.

2. La estimarea lui $\bar{\bar{x}}$ după datele obținute la un eșantion.

Scanned with OKEN Scanner

Exemplul 2:

La un eșantion s-au obținut: $\bar{x} \pm s = 30 \pm 10$ din $n = 50$ elemente.

Să se estimeze ce rezultat ar fi fost la întreaga populație.

$$\bar{\bar{x}} = \bar{x} \pm \sigma_{\bar{x}} = 30 \pm \frac{10}{\sqrt{50-1}} = \frac{10}{\sqrt{49}} = \frac{10}{7} = 1{,}19$$

Rezultatul este: $\bar{\bar{x}} \pm \sigma_{\bar{x}} = 30 \pm 1{,}19$

Unii autori mai „severi" cînd au un singur eșantion, întrebuințează formula: $ES = \frac{s}{\sqrt{n}} \cdot \left(1 - \frac{n}{N}\right)$, unde N=numărul elementelor din întreaga populație.

3. La estimarea lui x din datele obținute la mai multe eșantioane.

Exemplul 3:

Dintr-o populație s-au obținut următoarele rezultate:

$\bar{x}_A \pm s_A = 20 \pm 15$ din $n = 30$ elemente
$\bar{x}_B \pm s_B = 30 \pm 15$ din $n = 30$ elemente
$\bar{x}_C \pm s_C = 40 \pm 15$ din $n = 30$ elemente

Să se estimeze $\bar{\bar{x}} \pm \sigma_{\bar{x}}$

$$\bar{\bar{x}} = (20 + 30 + 40) : 3 = 30$$

Deoarece s-urile sînt egale și n-urile sînt egale:

$$\sigma_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{15}{\sqrt{3}} = \frac{15}{1{,}73} = 8{,}6$$

Rezultatul este: $\bar{\bar{x}} \pm \sigma_{\bar{x}} = 30 \pm 8{,}6$

Observație: $\sqrt{n} = \sqrt{3}$, deoarece am luat 3 valori s în calcul. Această metodă se întrebuințează cînd distribuția este sigur gaussiană.

Exemplul 4:

Dintr-o populație s-au obținut următoarele rezultate:
$\bar{x}_A \pm s_A = 20 \pm 10$ din 30 elemente
$\bar{x}_B \pm s_B = 30 \pm 15$ din 30 elemente
$\bar{x}_C \pm s_C = 40 \pm 20$ din 30 elemente

Scanned with OKEN Scanner

Să se estimeze $\bar{\bar{x}} \pm \sigma_{\bar{x}}$

$$\bar{\bar{x}} = (20+30+40) : 3 = 30$$

$$s = \sqrt{\frac{(\bar{x}_A - \bar{\bar{x}})^2 + (\bar{x}_B - \bar{x})^2 + (\bar{x}_C - \bar{\bar{x}})^2}{n=1}} =$$

$$\sqrt{\frac{(20-30)^2 + (30-30)^2 + (40-30)^2}{3-1}} = \sqrt{\frac{10^2+10^2}{2}} = \sqrt{\frac{200}{2}} =$$

$$= \sqrt{100} = 10$$

$$\sigma_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{10}{\sqrt{3}} = 5{,}7$$

Rezultatul este: $\bar{\bar{x}} \pm \sigma_{\bar{x}} = 30 \pm 5{,}7$

Această metodă se întrebuințează cînd populația nu este sigur gaussiană. Problema se poate rezolva însă și astfel:

n	x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
1	20	20—30=10	100
2	30	30—30= 0	0
3	40	40—30=10	100
3	90		200

$$\bar{x} = \sqrt{\frac{(x-\bar{x})^2}{n(n-1)}} = \sqrt{\frac{200}{3{,}2}} = \sqrt{\frac{200}{6}} = \sqrt{33{,}3} = 5{,}7$$

Deci rezultatul este egal după ambele metode.

Exemplul 5:

Dintr-o populație s-au obținut următoarele valori:

$\bar{x}_A \pm s_A = 20 \pm 10$ din $n = 25$ elemente
$\bar{x}_B \pm s_B = 30 \pm 20$ din $n = 35$ elemente
$\bar{x}_C \pm s_C = 40 \pm 30$ din $n = 45$ elemente

Să se estimeze $\bar{\bar{x}} \pm \sigma_{\bar{x}}$

$$\bar{x} = \frac{(20 \cdot 25) + (30 \cdot 35) + (40 \cdot 45)}{25+35+45} = \frac{500 + 1\,050 + 1\,800}{105} = \frac{3\,350}{105} = 31{,}9$$

Scanned with OKEN Scanner

Rezultatul este: $\bar{\bar{x}}=31,9$, deci s-a obținut o medie aritmetică ponderată din mediile eșantioanelor. Nu s-a putut calcula $\sigma_{\bar{x}}$ din abaterile eșantioanelor, deoarece n-urile nu sînt egale și s-urile nu sînt egale.

Exemplul 6:

Dintr-o populație cu $N=1\,000$ elemente, dintr-un eșantion cu $n=200$ elemente, s-a obținut $\bar{x}\pm s=100\pm 20$. Să se estimeze $\bar{\bar{x}}\pm\sigma_{\bar{x}}$.

Formula ce se utilizează este: $\bar{\bar{x}}=\bar{x}\pm(ES\cdot P\sigma)$, unde:

„$P\sigma$" = probabilitatea σ.

Calculul se face astfel: $ES=\frac{s}{\sqrt{n}}=\frac{20}{\sqrt{200}}=\frac{20}{14,1}=1,41$

Valoarea „$P\sigma$" o luăm din tabelul cu valorile curbei Gauss, avînd următoarele valori:

1 ES=66 % probabil=0,97 σ (căci 0,33 ·2=0,66 = =66 %)

2 ES=95,5% probabil=2 σ (căci 0,477·2=0,954= =95,5%)

3 ES=98 % probabil=2,33 σ (căci 0,49 ·2=0,98 = =98 %)

4 ES=99 % probabil=2,5 σ (căci 0,493·2=0,987= =99 %)

Deci: $\bar{\bar{x}}=\bar{x}\pm(ES\cdot P\sigma)=100\pm(1,41\cdot 0,97\ \sigma)=100\pm 1,36$.

Comentarii:

1. ES se utilizează pentru a compara rezultatele obținute la eșantioane, în ceea ce privește precizia.

2. Din rezultatele obținute la cîteva eșantioane, putem obține o „medie a mediilor" (medie aritmetică simplă sau ponderată), dar din s-urile lor nu putem obține o medie. Cu ajutorul lui ES obținem „un fel de abatere" a mai multor medii $\bar{x}$, așa că la o medie $\bar{\bar{x}}$ „sigură" îi adăugăm o „abatere orientativă", o „sigma artificială" valabilă pentru întreaga populație, deci nu o „abatere standard" σ, ci o „eroare standard" $\sigma_{\bar{x}}$, deoarece provine din calculare cu ajutorul $\bar{x}$-urilor. Deci cu ajutorul ES „extrapolăm" la întreaga populație ce am găsit la eșantion.

Scanned with OKEN Scanner

3.4.2. Eroarea Probabilă EP sau x_1

Unii autori consideră că estimarea lui $\bar{\bar{x}}$ se face mai corect cu ajutorul lui EP a cărui formulă este: EP= =ES·0,6745.

Explicație: 0,6745 σ=0,2486 din suprafață ≈25% din totalul suprafeței de sub curba lui Gauss.

Exemplul 1:

La un eșantion s-a găsit $\bar{x} \pm s = 20 \pm 15$ din n=25 elemente.

Să se estimeze $\bar{\bar{x}}$ cu ajutorul lui EP.

$$EP = \frac{15}{\sqrt{25}} = \frac{15}{5} = 3$$

$$EP = 3 \cdot 0{,}6745 = 2$$

Rezultatul este: $\bar{\bar{x}} \pm \sigma_{\bar{x}} = 20 \pm 2$.

3.4.3. Logica utilizării valorii ES și EP

Logica utilizării valorii ES (și EP) este următoarea: Ținînd seama de faptul că în calcule nu putem obține decît rezultate apropiate de „exactul absolut", ne putem stabili pretențiile în ceea ce privește proporția greșelilor ce le acceptăm. Astfel, dacă admitem ca din două rezultate, unul să fie greșit, înseamnă că avem o probabilitate de 50%; dacă acceptăm ca din 3 rezultate unul să fie greșit, avem o „siguranță" de 66% etc. În tabelul următor sînt redate valorile cele mai obișnuite admise în calcule, valoarea sigma fiind determinată de suprafața de sub curba lui Gauss, de ex. ±2 sigma=2·0,4772=0,9544≈ ≈95,5% probabilitate.

GREȘELI/ DETERMINĂRI	SIGURANTA PROBABILĂ	±σ	ES	EP
1/2	50%	0,66 σ	—	1
1/3	66	0,97	1	—
1/5	80	1,25	—	2
1/10	90	1,60	—	—

		±σ		
1/20	95	1,96	—	3
1/22	95,5	2,00	2	—
1/25	96	2,05	—	—
1/50	98	2,33	3	—
1/100	99	2,57	4	4

3.5. Testul t

STUDENT (pseudonimul lui Gosset) a calculat valorile lui t după formula: $t=\frac{(\bar{x}-\bar{\bar{x}})}{ES}$, de unde: $\bar{\bar{x}}=\bar{x}\pm(ES\cdot t)$. De aici rezultă că putem utiliza testul t în următoarele cazuri:

a) estimarea (stabilirea cu aproximație) a lui $\bar{\bar{x}}$ cînd avem ceilalți parametri ai eșantionului, utilizînd tabelul nr. 2 (cînd elementele provin dintr-un singur șir),

b) compararea a două medii, utilizînd tabelul nr. 4 (cînd valorile provin din două șiruri de valori). Această comparație se tratează la capitolul: Compararea mediilor.

Testul t se utilizează cînd numărul elementelor este mic, deci distribuția nu este gaussiană, astfel că puterea probantă a lui σ și s este redusă. Totuși, numărul elementelor nu trebuie să fie sub 5; de preferat mai multe (în lucrare utilizăm valori puține pentru a nu deveni plictisitoare calculele).

După efectuarea calculelor, se face interpretarea rezultatului, după tabelul 2, astfel:

— cînd $t_{calculat} > t_{tabel}$: coeficientul de corelație este semnificativ;

— cînd $t_{calculat} < t_{tabel}$: coeficientul de corelație este nesemnificativ, legătura dintre caracteristici este întîmplătoare.

Gradele de libertate se iau: n—1.

Exemplul 1:

La o analiză s-au întrebuințat următoarele cantități de soluție: 3 ml; 3 ml; 4 ml; 5 ml; 8 ml; 10 ml. Întrebarea este: aceste cantități au fost absolut necesare? Nu este vreo greșeală? (ținînd seama că există mari diferențe între valori). Deci dorim să estimăm pe $\bar{\bar{x}}$, pentru a stabili ce cantități de soluție ar fi necesare dacă s-ar efectua mai multe analize de acest fel.

Pentru rezolvare efectuăm următoarele calcule:

n	x_i	$(x_i-\bar{x})$	$(x_i-\bar{x})^2$
1	3	3—5,5=2,5	6,25
2	3	3—5,5=2,5	6,25
3	4	4—5,5=1,5	2,25
4	5	5—5,5=0,5	0,25
5	8	8—5,5=2,5	6,25
6	10	10—5,5=4,5	20,25
6	33		41,5

$$\bar{x}=33:6=5.5$$

$$s=\sqrt{\frac{41,5}{6-1}}=\sqrt{\frac{41,5}{5}}=2,9$$

Fiind valori puține, distribuția (populația) nu este „tocmai" gaussiană, deci s=2,9 nu are mare putere probantă. De altfel, 1 s înseamnă doar 68% siguranță, astfel că trebuie să calculăm:

$$ES=\frac{s}{\sqrt{n}}=\frac{2,9}{\sqrt{6}}=1,18$$

Calculînd pe ES cu varianta a doua, avem:

$$ES=\sqrt{\frac{(x-\bar{x})^2}{n\cdot(n-1)}}=\sqrt{\frac{41,5}{6,5}}=\sqrt{\frac{41,5}{30}}=\sqrt{1,38}=1,18$$

Deci am obținut același rezultat (1,18) cu ambele formule.

De aici avem: $\bar{\bar{x}}_t=\bar{x}\pm(ES\cdot t)=5,5\pm(1,18\cdot t)$

În tabelul nr. 2 la 6—1=5 grade de libertate (deoarece n=6) avem:

2 la p=0,1 =90% probabilitate
2,5 la p=0,05=95% probabilitate
4 la p=0,01=99% probabilitate

Dorind să ne asigurăm pentru o probabilitate de 95%, calculul este: $\bar{\bar{x}}_t=5,5\pm(2,5\cdot1,18)=5,5\pm3$
Rezultatul se exprimă: $\bar{\bar{x}}_t=5,5\pm3$ la p=0,05

Aceasta înseamnă: $\bar{\bar{x}}$ estimat cu testul t poate avea valori de 2,5 . . . 8,5 în 95% din cazuri, fără ca aceasta să însemne greșeală de tehnică.

Exemplul 2:

Un elev a fost testat (prin răspunsuri, extemporale etc.) și a obținut următoarele note: 6; 7; 8; 9. Întrebarea este: dacă elevul ar fi fost testat în continuare, ce note ar fi obținut? Cu alte cuvinte, plecînd de la un eșantion ($\bar{x}$) să estimăm întreaga populație ($\bar{\bar{x}}$).

Calculele sînt următoarele:

$$\bar{x}=(6+7+8+9):4=7,5$$

s=1,3 (după calculul obișnuit)

$$ES=\frac{s}{\sqrt{n-1}}=\frac{1,3}{\sqrt{3}}=0,75$$

Aplicarea testului t se face astfel:
grade de libertate=4 (note)—1=3
La acest șir găsim la 0,05 (=95%) cifra 3,1

$$\bar{\bar{x}}_t=7,5\pm(0,75\cdot3,1)=7,5\pm2,3$$

Rezultatul se exprimă: $\bar{\bar{x}}_t=7,5\pm2,3$ la p=0,05

Aceasta înseamnă că elevul va obține note de 7,5±±2,3 în 95% din cazuri. Nu ar fi fost potrivit dacă am fi dorit să ne asigurăm pentru o probabilitate de 99%, căci atunci, după calcule, elevul ar fi obținut note de 7,5±4,38, adică ar obține note între 3,12 și 11,88. Deci cu cît este mai largă „plaja" de asigurare, cu atît aplicarea testului este mai nesemnificativă (căci în orice caz, elevul ar obține note de 1 . . . 10).

Exemplul 3:

Într-un număr de 50 rezecții de stomac, au decedat 3 persoane. Întrebarea este: această proporție este absolut sigură? Sau cu alte cuvinte: la un număr mai mare de operații se va menține aceeași proporție de decese?

Rezolvarea se face astfel: $\bar{x}=6\%$ (3 decese la 50 cazuri=6 la 100).

$$s=\sqrt{\frac{p\cdot(100-p)}{n}}=\sqrt{\frac{6\cdot(100-6)}{50}}=\sqrt{\frac{6{,}94}{50}}=$$

$$=\sqrt{\frac{564}{50}}=3{,}4$$

Deci avem: $\bar{x}\pm s=6\pm3{,}4\%$

$$ES=\frac{3{,}4}{\sqrt{50}}=\frac{3{,}4}{7{,}07}=0{,}48$$

$$\bar{\bar{x}}_t=\bar{x}\pm(ES\cdot t)=6\pm(0{,}48\cdot t)$$

t la 50—1=49 grade libertate și p=0,05 (95%)=2
Rezultatul este: $\bar{\bar{x}}_t=6\pm(0{,}48\cdot2)=6\pm0{,}96$ la p=0,05.

Observații:

p=procentajul

Cînd numărul n este mic, se utilizează în calcul n—1.

Prin metode de calcul asemănătoare, s-a putut stabili că siguranța de funcționare a modulului lunar Apollo 11, care a coborît pe Lună la 20 aprilie 1969, era de 99,996%. Modulul avea 2 milioane de piese (față de cîteva mii ale unui automobil), cu 40 indicatoare, 71 becuri de control și 506 comutatoare.

3.6. Testul x^2

Făcînd comparație între caracteristicile a două eșantioane (ce provin din aceeași populație sau din populații diferite), găsim diferențe între ele. Întrebarea este: sînt importante aceste diferențe? Sînt semnificative? Cît de semnificative?

De cele mai multe ori avem caracteristicile unui eșantion pe care le comparăm cu caracteristicile unui eșantion teoretic, ipotetic, întrebarea fiind aceeași: diferențele sînt semnificative? Rezultatele obținute sînt în concordanță cu cele prevăzute teoretic? Stabilirea acestor diferențe (neconcordanțe) se face cu ajutorul testului χ^2, rezultatul exprimîndu-se printr-o cifră. În principiu, cu cît această cifră este mai mare, cu atît diferența între eșantioane este mai mare. Cînd diferența este mică, aceasta se poate datora întîmplării, de fapt diferența putînd fi (aproape) nulă, caz în care rezultatul se interpretează: se acceptă ipoteza nulă în sensul că diferența dintre eșantioane este (aproape) nulă. Cînd există diferență mai mare între eșantioane, se exprimă: se respinge ipoteza nulă, adică diferența nu este nulă, motiv pentru care se spune: diferența este semnificativă.

Exemplu:

Rezultatul calculului a dat cifra 0,0158. Această cifră o găsim în tabelul χ^2 la p=0,9. Aceasta înseamnă că eșantionul real se deosebește de cel teoretic în 1—0,9=10% din cazuri. Deci dacă avem 100 astfel de eșantioane, 10 din ele ar putea fi deosebite, restul fiind foarte asemănătoare cu eșantionul comparat. Deci diferența este nesemnificativă. Dacă cifra rezultată este 6,63, pe care o găsim în tabel la p=0,01, înseamnă că în 1—0,01=0,99=99% din cazuri există discordanță între eșantioanele comparate. Cu alte cuvinte, diferența este semnificativă cu o probabilitate de 99%. Metoda nu ne poate spune care din rezultate este mai bun, mai precis (cel observat sau cel teoretic), pentru aceasta urmînd a utiliza alte metode de cercetare.

Deci testul χ^2 se utilizează pentru:

— a stabili diferența, discordanța, dintre două eșantioane, ajutîndu-ne să respingem ipoteza neverosimilă că eșantionul A este egal cu eșantionul B, cînd de fapt nu sînt egale. Din această cauză se numește și „test de verosimilitate";

— a stabili dacă între două caracteristici (două eșantioane) există vreo concordanță, sau sînt diferite și independente una de alta. Din această cauză se mai numește și „test de independență";

— a stabili dacă două eșantioane provin de la aceeași populație „normală", omogenă. Din această cauză se numește și „test de omogenitate";

— a stabili dacă între cele două eșantioane există discordanță sau concordanță, numindu-se deci și „test de concordanță";

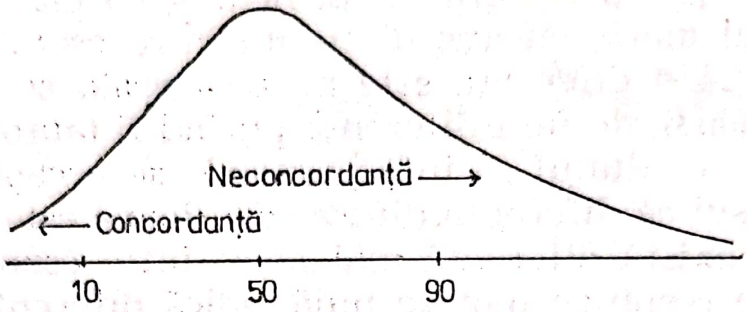


FIG. 21.

— și în sfîrșit, pentru toate aceste motive, se numește și „test de probabilitate exactă".

Grafic, repartiția χ^2 este reprezentată în fig. 21, unde se vede că pe măsură ce scade concordanța, crește neconcordanța (suprafața de sub curbă rămînînd 100%).

Formula sa este: $\chi^2 = \Sigma \frac{(FO - FT)^2}{FT}$, unde:

FO = Frecvența Observată, adică rezultatul obținut prin măsurări, cîntăriri, titrări etc.

FT = Frecvența Teoretică (așteptată, ipotetică sau apreciată prin formule matematice sau metode experimentale prealabile).

Deoarece la numărător avem valoarea la patrat, testul se numește „patrat", spre deosebire de alți termeni asemănători întrebuințați în statistică (χ, χ_1 etc.).

Datorită multiplelor întrebuințări, calculele valorii χ^2 au fost tabelate (tabel nr. 3) în funcție de diferite probabilități.

Observații:

1. Elementele eșantioanelor trebuie să fie independente, pentru ca deducția să fie valabilă;

2. Cînd frecvența teoretică (așteptată, ipotetică) este mică, se pot obține rezultate eronate;

3. Cînd una din frecvențe este sub 5, valoarea respectivă se alătură la clasa (coloana) vecină, așa ca să fie cel puțin 10 elemente;

4. Numărul gradelor de libertate (gr. lib.) se stabilește după principiul: nr. elemente independente — nr. elemente dependente. În general, gradul de libertate=(nr. coloane—1) (nr. șiruri—1), deci trebuie să fie cel puțin 2 coloane și 2 șiruri;

5. Cînd există un singur grad de libertate, e de preferat a exista cel puțin 10 elemente;

6. Testul χ^2 se aplică numai cînd populația este gaussiană;

7. Cînd avem un număr redus de valori (elemente) și deci populația nu este gaussiană, cercetarea se face mai bine cu ajutorul testului t, căci χ^2 cu 1 grad de libertate corespunde cu testul t cu ∞ grade de libertate.

Exemplul 1:

Un elev a obținut următoarele note: 5; 5; 6; 7. El susține că trebuia să primească note de 10. Să se stabilească cu ajutorul testului χ^2 concordanța dintre eșantioane (note obținute și note așteptate).

Deoarece problema este simplă, întocmim un singur tabel:

F.O.	F.T.
5	10
5	10
6	10
7	10

Calculul este următorul: $\chi^2=\Sigma\frac{(F\,O-F\,T)^2}{F\,T}=\frac{(5-10)^2}{10}+$

$$+\frac{(5-10)^2}{10}+\frac{(6-10)^2}{10}+\frac{(7-10)^2}{10}=\frac{25}{10}+\frac{25}{10}+\frac{16}{10}+\frac{9}{10}=$$

$$=2{,}5+2{,}5+1{,}6+0{,}9=7{,}5$$

Gr. lib.=(2 col—1)·(4 șir—1)=1·3=3

La 3 gr. lib., cifra 7,5 se găsește la p=0,05 (de fapt 7,8). Rezultatul se exprimă: χ^2=semnificativ (discordant) la p=0,05.

Exemplul 2:

Să presupunem cp pe o lamă hemocitometrică, elementele (globulele roșii) s-au repartizat uniform cîte 10 elemente pe fiecare pătrat. Să se verifice verosimilatea situației. În acest scop întocmim tabelul:

ŞIR	PATRAT	COL 1 F.O.	COL 2 F.T.	COL 3 (F.O.—F.T.)²
1	1	10	10	$(10-10)^2=0$
2	2	10	10	$(10-10)^2=0$
3	3	10	10	$(10-10)^2=0$
4	4	10	10	$(10-10)^2=0$

Calculul se face astfel: $\chi^2=\Sigma\frac{FO-FT}{FT}=\frac{0}{10}+\frac{0}{10}+\frac{0}{10}+\frac{0}{10}=0$

Gr. lib.=(2 col—1)·(4 şir—1)=1·3=3

Interpretare: Rezultatul 0 arată că nu este nici o discordanță între eșantioane. De fapt, la 3 gr. lib. cifra cea mai mică (0,58) o găsim la p=0,90. Deci eșantioanele concordă în peste 90% din cazuri. În realitate nu se întîmplă așa, căci elementele nu se înșiră disciplinat, ordonat, uniform, ca soldații, pe patrate, ci ca în exemplul următor.

Exemplul 3:

Pe o lamă hemocitometrică 40 elemente s-au repartizat astfel (coloana 1 din tabel). După părerea noastră ar fi trebuit să se distribuie uniform (ca în coloana 2).

ŞIR	PATRAT	F.O. COL 1	F.T. COL 2	(F.O.—F.T.)² COL 3
1	1	2	10	$8^2=64$
2	2	12	10	$2^2=4$
3	3	18	10	$8^2=64$
4	4	8	10	$2^2=4$

Scanned with OKEN Scanner

De aici calculăm:

$$\chi^2=\frac{64}{10}+\frac{4}{10}+\frac{64}{10}+\frac{4}{10}=6{,}4+0{,}4+6{,}4+0{,}4=13{,}6$$

$$\text{Gr. lib.}=(2\ \text{col}-1)\cdot(4\ \text{șir}-1)=1\cdot 3=3$$

Cifra 13,6 o găsim la 3 gr. lib. la p=0,01, deci eșantioanele se deosebesc foarte mult. Dar s-a „strecurat" o greșeală: în patratul 1 au fost numai 2 elemente (deci prea puține), pe care trebuie să le alăturăm șirului 2 din aceeași coloană, deci avem:

ȘIR	PATRAT	F.O.	F.T.	(F.O.—F.T.)²
		COL 1	COL 2	COL 3
1	1	14	20	$6^2=36$
2	2	18	10	$8^2=64$
3	3	8	10	$2^2=4$

De aici calculăm: $\chi^2=\frac{36}{20}+\frac{64}{10}+\frac{4}{10}=1{,}8+6{,}4+0{,}4=8{,}6$

$$\text{Gr. lib.}=(2\ \text{col}-1)\cdot(3\ \text{șir}-1)=1\cdot 2=2$$

Cifra 8,6 la 2 gr. lib. se află la p=0,01 (de fapt 9,2), ceea ce înseamnă că χ^2=semnificativ (discordant) la p=0,01 (=99% din cazuri). Discordanța se datorește dorinței noastre ca totul să se aranjeze ca în coloana 2.

Exemplul 4:

Dintr-un lot de 60 animale de experiență, au decedat 33 animale, iar din lotul martor au decedat 6 animale. Întrebarea este: această mortalitate este întîmplătoare sau nu?

Se întocmește tabelul:

		F.O.			
		COL 1	COL 2	COL 3	OBSERVAȚII
		DECEDAT	NU AU DECEDAT	TOTAL	
șir	1	33	27	60	lot experim.
șir	2	6	24	30	lot martor
șir	3	39	51	90	total

Scanned with OKEN Scanner

La această experiență nu am putut prevedea nici un rezultat, deci nu avem o frecvență teoretică sau așteptată. Deci trebuie să ne „confecționăm" o astfel de frecvență, calculînd ce valori ar fi fost dacă totul se desfășura întîmplător.

În acest scop efectuăm un calcul după modelul din tabel:

ŞIR	COL 1	COL 2	TOTAL
1	a	b	n_1
2	c	d	n_2
total	n_3	n_4	N

Calculul se face astfel:

$$\frac{n_1 \cdot n_3}{N};\ \frac{n_1 \cdot n_4}{N};\ \frac{n_2 \cdot n_3}{N};\ \frac{n_2 \cdot n_4}{N}$$

Obținem F.T. astfel:

	F.T.			
ŞIR	COL 1	COL 2	COL 3	OBSERVAȚII
1	$\frac{66,39}{90}=26$	$\frac{60,51}{90}=34$	60	lot exper.
2	$\frac{30,39}{90}=13$	$\frac{30,51}{90}=17$	30	lot martor

Cu aceste cifre teoretice, calculăm:

COL 1	COL 2	OBSERVAȚII
$\frac{(33-26)^2}{26}=\frac{7^2}{26}=1,88$	$\frac{(27-34)^2}{34}=\frac{-7^2}{34}=1,44$	exper.
$\frac{(16-13)^2}{13}=\frac{-7^2}{13}=3,76$	$\frac{(25-17)^2}{12}=\frac{-7^2}{17}=2,88$	mart.

Rezultatul este: $\chi^2=1,88+1,44+3,76+2,88=9,96$

Gr. lib.$=(2$ col$-1)\cdot(2$ şir$-1)=1\cdot1=1$

Scanned with OKEN Scanner

Cifra 9,96 la 1 gr. lib. se află la p=0,001 și rezultatul se exprimă: χ^2 este semnificativ (discordant) la p=0,001.

La problema respectivă putem face următoarea apreciere:

la 60 animale au decedat 33=55%

la 30 animale au decedat 6=20%, dar în alte cazuri mai complicate, nu putem stabili deosebirea decît cu testul χ^2.

Exemplul 5:

Un număr de 308 bolnavi au fost tratați cu 3 medicamente diferite (sau în doze diferite). În acest scop au fost împărțiți în 3 eșantioane:

eșantionul A=100 pacienți, tratați cu medicamentul a
eșantionul B=102 pacienți, tratați cu medicamentul b
eșantionul C=106 pacienți, tratați cu medicamentul c.

Deoarece se consideră că medicamentul a este cel mai bun (sau administrat în doză mare), ar fi trebuit ca eșantionul A să aibă numai rezultate foarte bune, la eșantionul B rezultate mijlocii, iar la eșantionul C rezultate slabe. Dar, s-au obținut următoarele rezultate:

eșantion A, tratament a: 50 foarte bine
35 bine
15 slab

eșantion B, tratament b: 30 foarte bine
44 bine
28 slab

eșantion C, tratament c: 12 foarte bine
24 bine
70 slab

Aceste rezultate le tabelăm astfel:

		F.O.			
ȘIR	EȘANTION	COL 1	COL 2	COL 3	TOTAL
1	A	50	35	15	100
2	B	30	44	28	102
3	C	12	24	70	106
4	Total	92	103	113	308

Scanned with OKEN Scanner

Deoarece nu avem o frecvență teoretică cu care să facem comparația, ne „confecționăm“ o astfel de frecvență, astfel:

ŞIR	COL 1	COL 2	COL 3	TOTAL
1	a	b	c	n_1
2	d	e	f	n_2
total	n_3	n_4	n_5	N

După acest tabel obținem valorile teoretice astfel:

$$\frac{n_1 \cdot n_3}{N};\ \frac{n_1 \cdot n_4}{N};\ \frac{n_1 \cdot n_5}{N};\ \frac{n_2 \cdot n_3}{N};\ \frac{n_2 \cdot n_4}{N};\ \frac{n_2 \cdot n_5}{N}$$

Calculele se efectuează astfel:

$$\frac{100 \cdot 92}{308} = 29{,}8 \qquad \frac{100 \cdot 103}{308} = 33{,}4 \qquad \frac{100 \cdot 113}{308} = 36{,}6$$

$$\frac{102 \cdot 92}{308} = 30{,}4 \qquad \frac{102 \cdot 103}{308} = 34{,}1 \qquad \frac{102 \cdot 173}{308} = 37{,}4$$

$$\frac{106 \cdot 92}{308} = 31{,}6 \qquad \frac{106 \cdot 103}{308} = 35{,}4 \qquad \frac{106 \cdot 113}{308} = 38{,}8$$

Pentru simplificare, așezăm rezultatele în tabel:

ŞIR	EŞAN-TION	COL 1	COL 2	COL 3	TOTAL
1	A	29,8	33,4	36,6	99,8
2	B	30,4	34,1	37,4	101,9
3	C	31,6	35,4	38,8	105,8
	total	91,8	102,9	112,8	307,5

Totalurile nu sînt absolut exacte (ex. 101,9 în loc de 102) din cauza neglijării unor zecimale.

Scanned with OKEN Scanner

După ce am obținut rezultatele valorilor teoretice, calculăm dacă există diferență întîmplătoare între eșantioane, astfel:

$$\chi^2 = \frac{(50-29,8)^2}{29,8} + \frac{(35-33,4)^2}{33,4} + \frac{(15-36,6)^2}{36,4} +$$

$$+ \frac{(30-30,4)^2}{30,4} + \frac{(44-34,1)^2}{34,1} + \frac{(28-37,4)^2}{37,4} +$$

$$+ \frac{(12-31,6)^2}{31,6} + \frac{(24-35,4)^2}{35,4} + \frac{(70-38,8)^2}{38,8} =$$

$$=13,69+0,07+12,81+0,05+2,87+2,36+12,56+$$
$$+3,67+25=73,03$$

Gr. lib.=(col—1)·(șir—1)=(3—1)·(3—1)=4

Interpretarea rezultatului:

a) Cea mai mare cifră la 4 gr. lib. este 18,5, deci diferența dintre eșantioane nu este întîmplătoare; eșantioanele se deosebesc foarte mult. Se exprimă: χ^2 este semnificativ (discordant) la peste 0,001. Cît și cum a influențat tratamentul nu se poate stabili cu metoda χ^2, urmînd a se face alte determinări: glicemie, greutate, suc gastric etc.

b) Interpretarea rezultatului se poate face și după o formulă orientativă (de ex. cînd nu avem la dispoziție tabel):

cînd χ^2 calculat $< (f+\sqrt{2f})$, diferența este întîmplătoare. În cazul nostru: $73,03 > (4+\sqrt{2\cdot 4}) = 73,03 > 6,8$, deci diferența nu este întîmplătoare, ci semnificativă.

c) O altă formulă orientativă este următoarea:

cînd $2 > (2\cdot\chi^2)-(2f-1)$, diferența este întîmplătoare. În cazul nostru: $(2\cdot 73,03)-(2\cdot 4-1)=144-8-1=9,35$, deci $2<9,35$, deci diferența nu este întîmplătoare.

În ambele formule, f=gr. de libertate.

În concluzie:

Textul χ^2 reprezintă, generalizat, o „distanță" între două fenomene, reprezentată în tabelul respectiv. Cu cît cifra obținută este mai mare, cu atît diferența este mai mare, iar la 0 înseamnă că eșantioanele sînt egale.

Metoda nu ne poate spune care din eșantioane este mai bun, mai corect, mai exact, ci doar că există sau nu diferență între ele.

Scanned with OKEN Scanner

Capitolul 4
Comparaţii

Uneori este nevoie de a face comparaţii între două sau mai multe valori obţinute prin calcule, pentru a vedea dacă între ele sînt diferenţe semnificative. Astfel de comparaţii se fac dacă se întrebuinţează aparate, metode sau reactivi diferiţi, stabilindu-se dacă precizia metodelor sau a aparatelor diferă între ele.

4.1. Compararea mediilor aritmetice

Exemplul 1:

Fie următoarele valori găsite la aceeaşi probă, dar analizele efectuate cu metode diferite:

$\bar{x}_1 \pm s_1 = 20 \pm 10$, provenind din $n=20$ elemente

$\bar{x}_2 \pm s_2 = 25 \pm 5$, provenind din $n=30$ elemente.

Să se compare mediile pentru a se vedea eventuale diferenţe între metodele întrebuinţate.

Formula este: $t = \dfrac{\bar{x}_1 - \bar{x}_2}{s' \cdot \sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$, calculul lui s' fiind:

$$s' = \sqrt{\frac{n_1 \cdot s_1^2 + n_2 \cdot s_2^2}{n_1 + n_2}} = \sqrt{\frac{20 \cdot 10^2 + 30 \cdot 5^2}{20+30}} = \sqrt{55} = 7{,}4$$

$$t = \frac{20-25}{7{,}4 \cdot \sqrt{\dfrac{1}{20} + \dfrac{1}{30}}} = \frac{5}{7{,}4 \cdot 0{,}28} = \frac{5}{2{,}07} = 2{,}07.$$

f. Grad de libertate $= (n_1 + n_2) - 2 = (20+30) - 2 = 48$ (cînd $s_1 \neq s_2$).

În tabelul t (pentru compararea mediilor) la f=48 cifra este 1,676 la p=0,95 (de fapt la f=50).

Deoarece $t_{calculat} > t_{tabel}$ adică 2,07>1,67, diferența este semnificativă la p=0,95. Deci mediile se deosebesc mult, fără a se putea preciza care medie este mai bună.

4.2. Compararea abaterilor patratice

Exemplul 1:

Fie următoarele valori obținute la două analize:
$x_1 \pm s_1 = 40 \pm 10$, provenind din n=20 elemente
$\bar{x}_2 \pm s_2 = 50 \pm 5$, provenind din n=10 elemente.
Să se stabilească dacă există diferențe între abateri.

După formula Fischer și Snedecor avem: $F = \frac{S_1^2}{s_2^2}$,

unde:

F=valoarea din tabelul lui Fischer
S=valoarea abaterii mai mari
s =valoarea abaterii mai mici

În cazul nostru: $F = \frac{10^2}{5^2} = \frac{100}{25} = 4$.

În tabelul Fisher, la f=20 elemente (cifra cea mai mare, citată pe abscisă), la intersecția cu f=10 elemente (cifra cea mai mică, citită pe ordonată) găsim cifra 2.77.

Deoarece $F_{calculat}$ 4>F_{tabel} 2,77, diferența dintre abateri este semnificativă la p=0,95.

Dacă citim în tabelul următor la f=20 elemente (pe abscisă) cu întretăierea cu f=10 elemente (pe ordonată), găsim cifra 4,41. Deoarece $F_{calculat}$ 4<F_{tabel} 4,41, diferența nu este semnificativă la p=0,99.

4.3. Coeficientul de variație

Trebuie ținut seama că nu se pot face comparații între două medii sau abateri decît dacă provin din eșantioane care au aceeași unitate de măsură. De ex. nu putem com-

para $\bar{x}$ kg cu $\bar{x}$ m. Totuşi se pot face comparaţii între diferite abateri, chiar dacă au unităţi de măsură diferite, obţinînd o abatere procentuală, sub denumirea de „coeficient de variaţie", formula fiind: $CV = \frac{s}{\bar{x}} \cdot 100$

Exemplul 1:

$$\bar{x}_1 \pm s_1 = 90 \pm 10 \text{ kg}$$

$$\bar{x}_2 \pm s_2 = 50 \pm 10 \text{ m}$$

Să se stabilească C V a acestor valori.

$$C V_1 = \frac{10}{90} \cdot 100 = 11\% \text{ (din } \bar{x}_1)$$

$$C V_2 = \frac{10}{50} \cdot 100 = 20\% \text{ (din } \bar{x}_2).$$

De aici rezultă că „populaţia" este mai omogenă la $\bar{x}_1$.

4.4. Indici

Indicii sînt valori ce arată un raport între diferite mărimi. Ei se exprimă în numere întregi, fracţii, procente etc. Indicii servesc pentru studierea evoluţiei unor fenomene;

— compararea fenomenelor de acelaşi fel, ce se desfăşoară în acelaşi timp, dar în spaţii diferite (sau invers);

— precizarea gradului de îndeplinire a planului de producţie.

4.4.1. Indicele simplu

Exemplul 1:

Într-o unitate s-au produs 120 piese în trim. I, 110 în trim. II şi 130 în trim. III.

Scanned with OKEN Scanner

Să se calculeze indicii de producţie comparativ cu trim. I.

Formula este: $\overline{I} = \frac{Nn}{N_0}$, unde:

I =indicele simplu
Nn=nivelul mediu la un moment dat
No=nivelul mediu la începutul seriei

În cazul de faţă:

$\overline{I}=1$ pentru trim. I

$\overline{I} = \frac{110}{120} = 0{,}91$ pentru trim. II

$\overline{I} = \frac{130}{120} = 1{,}08$ pentru trim. III.

4.4.2. Indicele procentual

Evoluţia fenomenului este mai evidentă dacă se calculează astfel:

Exemplul 1:

$\overline{I}=100\%$ pentru trim. I

$\overline{I} = \frac{110}{120} \cdot 100 = 91\%$ pentru trim. II

$\overline{I} = \frac{130}{120} \cdot 100 = 108\%$ pentru trim. III

4.4.3. Indicele cronologic

Exemplul 2:

Într-un spital s-au internat 100 bolnavi în ianuarie, 110 în februarie, 115 în martie.

Care este nivelul mediu $\overline{N}$ al termenilor seriei, sau cu alte cuvinte, care este media lunară a bolnavilor internaţi?

$$\overline{N} = \overline{x}_{cron} = \frac{\frac{100}{2} + 110 + \frac{115}{2}}{3-1} = \frac{275}{2} = 108{,}75$$

Scanned with OKEN Scanner

4.4.4. Nivelul mediu procentual

Exemplul 3:

În luna ianuarie planul a fost îndeplinit 106%, în februarie 106%, iar în restul lunilor 100%. Care este nivelul mediu procentual?

$$\overline{N}\% = \overline{x}_{geom\ s} = \sqrt[n]{i_1 \cdot i_2 \ldots i_n} = \sqrt[12]{106.106.100.100 \ldots 100},$$

ceea ce se rezolvă astfel:

$$\log \overline{N}\% = \frac{(2 \cdot \log 106) + (10 \cdot \log 100)}{12} = \frac{2 \cdot 2{,}02531) + (10 \cdot 2)}{12} =$$

$$= \frac{(4{,}05062) + (20)}{12} = \frac{24{,}05062}{12} = 2{,}004218$$

Antilogaritmul mantisei 004218=101, deci

$$\overline{N}\% = 101\%$$

Ritmul mediu procentual

$$\overline{R}\% = \overline{N}\% - 100 = 101 - 100 = +1\% \text{ anual.}$$

În concluzie:

Indicele procentual demnostrează mai evident evoluţia fenomenelor; pentru valorile ce evoluează în timp, se utilizează $\overline{x}_{cron\ s}$; pentru valorile ce cresc sau scad în timp, se utilizează $\overline{x}_{geom}$.

4.4.5. Entropia informaţională

Asemănător cu entropia termodinamică stabilită de Clausius, Shanon a stabilit şi o entropie informaţională, care prelucrată de Liteanu şi Crişan are următoarea formulă: $EI \approx \log \frac{\overline{x}}{0{,}8\ s}$

Exemplul 4:

Prin titrimetrie am găsit $x_1 \pm s_1 = 10 \pm 5$

Prin gravimetrie am găsit $x_2 \pm s_2 = 10 \pm 4$.

Scanned with OKEN Scanner

Să se stabilească gradul de informație ce se obține de la ambele valori.

$$E\,I_1 = \log \frac{10}{0{,}8 \cdot 5} = \log \frac{10}{4} = \log 2{,}5 = 0{,}3979$$

$$E\,I_2 = \log \frac{10}{0{,}8 \cdot 4} = \log \frac{10}{3{,}2} = \log 3{,}125 = 0{,}4949$$

Deci obținem informații mai ample prin metoda gravimetrică.

Dacă antilogaritmăm obținem: 2,50 la titrimetrie și 3,126 la gravimetrie.

Scanned with OKEN Scanner

Capitolul 5

Eliminarea valorilor eronate

Eliminarea valorilor eronate sau afectate de erori grosolane față de șirul de valori reale trebuie făcută pentru ca acestea să nu afecteze rezultatele (medii, indici, grafice). Erorile pot apare ca greșeli de citire, de transcriere, de calcul. Erorile susceptibile a fi eronate sînt cele de la extremitățile șirului de valori (acestea fiind așezate în șir crescător sau descrescător). Verificarea și eliminarea lor se face prin una sau mai multe metode și anume:

1. testul celor 3 sigme
2. „ Grubbs
3. „ Q (Dean și Dixon)
4. „ t
5. „ Romanovski
6. „ λ (Irwin)

Pentru exemplificare, redăm aceeași problemă soluționată prin metodele menționate.

Exemplul 1:

Fie următoarele valori: 2; 52; 50; 54. Să se stabilească dacă valoarea dubioasă $x_d=2$ este eronată față de șirul de valori și dacă trebuie eliminată din calcule (fiind de 27 de ori mai mică decît valoarea cea mai mare). Pentru soluționare, ordonăm șirul în ordine crescîndă: 2; 50; 52; 54.

5.1. Testul celor „3 sigme"

Metoda constă în cercetarea dacă $\frac{x_d-\bar{x}}{s}<3$. Logica este următoarea: probabilitatea ca un eveniment contrar să aibă valori în afara intervalului $\pm 3\sigma$ (ceea ce reprezintă 99,76% din totalul valorilor) este foarte mică (100—99,76 =

$=0{,}26\% = 2{,}6\%$), deci practic un astfel de eveniment poate fi considerat (aproape) imposibil.

Pentru a aplica formula trebuie calculat $\bar{x} \pm s$:

n	x_i	x_i^2
1	2	4
2	50	2 500
3	52	2 704
4	54	2 916
4	158	8 124

$$\bar{x} = 158:4 = 39{,}5$$

$$s = \sqrt{\frac{8\,124 - \frac{158^2}{4}}{4-1}} = \sqrt{\frac{8\,124 - 6\,241}{3}} = \sqrt{627} = 25$$

Rezultatul este: $\bar{x} \pm s = 39{,}5 \pm 25$

Aplicînd formula avem: $\frac{2-39{,}5}{25} = 1{,}5$

Răspunsul este: deoarece $1{,}5 < 3$ (sigme), valoarea $x_d = 2$ nu trebuie eliminată.

5.2. Testul Grubbs

Formula este: $G = \frac{x_d - \bar{x}}{s}$, unde:

G = valoarea din tabelul Grubbs.

În cazul de față: $G = \frac{2-39{,}5}{25} = 1{,}5$

G_{tabel} la 4 valori și $p = 95\% = 1{,}77$.

Răspunsul este: deoarece G_{tabel} $1{,}71 > G_{calculat}$ $1{,}5$ nu se elimină, valoarea $x_d = 2$. După cum se vede, este o variantă a testului „3 sigme".

5.3. Testul Q (Dean și Dixon)

Formula este: $Q = \frac{x_d - x_a}{x_d - x_m}$, unde:

x_a = valoarea cea mai apropiată
x_m = valoarea cea mai mică din serie
Q = valoarea din tabel

În cazul de față: $Q = \frac{2-50}{2-54} = 0{,}92$

Q_{tabel} la 4 valori și $p = 95\% = 0{,}77$

Răspunsul este: deoarece Q_{tabel} $0{,}77 > Q_{calculat}$ $0{,}92$, valoarea x_d se elimină.

5.4. Testul t

Formula este: $t = \frac{x_d - \bar{x}}{s}$

În cazul nostru: $t = \frac{2-39{,}5}{25} = 1{,}5$

t_{tabel} la $4-1=3$ gr. libertate și $p = 95\% = 3{,}18$

Răspunsul este: deoarece $t_{calculat}$ $1{,}5 < t_{tabel}$ $3{,}18$, nu se elimină x_d.

Observație: se întrebuințează tabelul t nr. 2, fiind un singur șir de valori.

5.5. Testul Romanovski

Formula este: $R = \frac{x_d - \bar{x}}{s' \sqrt{\frac{n}{n-1}}}$, unde:

s' = abaterea calculată fără x_d
$n-1$ = șirul de valori fără x_d

Scanned with OKEN Scanner

Aplicînd formula, avem:

n	x_i	$x_i-\bar{x}$	$(x_i-\bar{x})^2$
1	2	—	—
2	50	2	4
3	52	0	0
4	54	2	4
	158		8

$$\bar{x}=158:4=39,5$$

$$s=\sqrt{\frac{8}{2}}=2$$

$$s'=s\sqrt{\frac{n}{n-1}}=2\sqrt{\frac{4}{3}}=2\cdot 1,15=2,3$$

De aici avem: $R=\frac{2-52}{2,3}=21,7$

Răspunsul este: deoarece $R_{calculat}$ 21,7 > R_{tabel} 3,56 (la f=4 și p=0,95) x_d se elimină. Deci testul Romanovski este un test mai sever decît testul t.

5.6. Testul λ (Irwin)

Formula este: $\lambda=\frac{x_d-x_a}{s}$, unde:

x_a=valoarea cea mai apropiată, valorile fiind așezate în șir crescător.

Utilizînd valorile calculate, avem:

$$\lambda=\frac{2-50}{25}=1,92.$$

λ_{tabel} la 4 valori și 95% probabilitate=2,05.

Răspunsul este: deoarece $\lambda_{calculat}$ 1,92 $<_{tabel}$ 2,05, valoarea x_d nu trebuie eliminată. Testul nu este prea eficient cînd x_d este destul de apropiat de $\bar{x}$ a șirului de valori.

5.7. Interpretarea rezultatelor

Tabelînd rezultatele obținute, am găsit că:

regula „3 sigme“	:	nu se elimină x_d
testul Grubbs	:	„
testul t	:	„
testul λ	:	„
testul Q	:	se elimină x_d
testul Romanovski	:	„

Deci după 4 teste nu se elimină, iar după 2 teste se elimină x_d. Pentru uniformizare, am luat p=95%, dar alt cercetător ar fi putut lua 90 sau 99%, deci eliminarea se face în funcție de exigența cercetătorului și formula aplicată (ținînd seama că practic nu se aplică toate acestea). În cazul dat, este de preferat a se elimina $x_d=2$, căci altfel înseamnă că acceptăm o mare dispersie a valorilor individuale $\bar{x}_i$, ceea ce se vede și din $\bar{x} \pm s = 39 \pm 25$. În principiu, valoarea x_d se elimină cînd valoarea calculată este mai mare decît valoarea din tabelul respectiv.

Observație: aceste metode se întrebuințează numai cînd valorile provin dintr-o distribuție gaussiană.

Scanned with OKEN Scanner

Capitolul 6

Eşantionul

Pentru a putea utiliza calculele statistice în studierea fenomenelor, trebuie respectate etapele cercetării statistice. Acestea sînt următoarele:

1. Observarea fenomenelor şi documentarea;
2. Alegerea eşantionului şi cercetarea sa, deci obţinerea informaţiilor;
3. Prelucrarea matematică a rezultatelor;
4. Întocmirea graficelor, tabelelor, rezultatelor;
5. Interpretarea rezultatelor, formularea concluziilor şi stabilirea unor prognoze şi măsuri.

În capitolul prezent, ne ocupăm de eşantion.

Pentru obţinerea de informaţii asupra fenomenului cercetat, studiem fenomenul pe toată întinderea sa: în timp, în spaţiu, pe număr total de locuitori sau pe un eşantion, pe sexe, pe vîrste etc.

Întrucît de cele mai multe ori nu dispunem de întreaga populaţie ne limităm la un număr mic de cazuri, luînd în cercetare numai un lot reprezentativ al mulţimii, adică un eşantion. După STAS 2631-72 prin eşantion se înţelege o submulţime sau o parte a populaţiei statistice, prelevată după unele procedee tehnice de sondaj, bine precizate. El serveşte drept sursă de informaţii pentru investigarea proprietăţilor întregii populaţii.

În statistică eşantionul se mai numeşte „selecţie", „subpopulaţie" sau „populaţie de sondaj". În industrie eşantionul reprezintă una sau mai multe din unităţile de produs, sau o cantitate de material dintr-un lot, în vederea determinărilor.

6.1. Alegerea eşantionului

Alegerea eşantionului are mare importanţă şi redăm ca exemplu următorul episod, care a consacrat în 1954

Scanned with OKEN Scanner

vaccinul antipoliomielitic Salk (vaccinul Sabin a fost introdus în 1959):

440 000 copii vaccinați cu vaccin Salk, 210 000 copii vaccinați cu excipient, 1 080 000 copii martori.

Rezultatele sînt redate în tabelul următor:

	CAZURI POLIO-MIELITĂ	PARA-LIZAȚI POLIO-MIELITĂ	CAZURI MOR-TALE	TOTAL
Vaccinați	113	71	1	440 000
Excipient	750	115	4	210 000
Martori		330	11	1 080 000
Total	863	516	16	1 730 000

Din tabel se vede că vaccinul a fost eficient atît la cazurile de poliomelită, cît și la cele cu paralizii și mortalitate. În S.U.A. au fost în 1935 un număr de 35 000 cazuri de poliomelită cu paralizii, iar după 10 ani de la aplicarea vaccinului numărul a scăzut la 121, ceea ce dovedește justețea aprecierii statistice asupra eficacității vaccinului, fiind justificați cei 7 ani de muncă, cele 30 000 de maimuțe sacrificate și cei 1 370 000 de dolari cheltuiți. Ca urmare a acestui vaccin, în 1955 boala a fost eradicată în Europa occidentală, S.U.A. etc. Un experiment asemănător s-a făcut cu ocazia testării eficacității unui vaccin antihepatitic B.

Principiile ce trebuie avute în vedere la alegerea eșantionului care reprezintă „o porțiune dintr-un univers", sînt următoarele:

1. Potrivit „legii numerelor mari", în general indicii sînt mai stabili și diferă mai puțin de indicii calculați la întreaga populație, cu cît eșantionul este mai mare. Cu cît populația este mai omogenă, cu atît eșantionul poate fi mai mic (număr de experiențe, măsurători, persoane etc.). În orice caz, eșantionul trebuie să conțină ceea ce este tipic, caracteristic, normal, obișnuit, pentru întreaga populație. Eșantionul trebuie să aibă și un martor, ales din aceeași grupare ca și eșantionul experimental. Lotul eșantion și lotul martor pot fi mai mici dacă legătura dintre

Scanned with OKEN Scanner

cauză și efect este mai evidentă, precum și dacă ambele loturi sînt omogene.

„Legea numerelor mari“ enunțată de Georg Mayr în 1899 în „Statistica și științele sociale“ constă în următoarele: în cazul numerelor mari, deci a populațiilor mari (d.p.d.v. statistic) reies anumite regularități și anume:

— regularitate în structura componentelor masei (=regularitatea componentelor);

— regularitatea în producerea anumitor fenomene (=regularitatea mișcării), lucruri ce pot fi constatate și prin cercetările cantităților mici, arbitrar alese, ale faptelor din care se însumează rezultatul general al observării de masă.

Matematicianul J. Bernoulli se exprimă astfel: dacă într-o colectivitate mare se ia un număr N suficient de mare de probe și din acestea n sînt favorabile, se poate admite probabilitatea că fenomenul cercetat, privind întreaga colectivitate, este dat de frecvența relativă f_n corespunzătoare formulei: $f_n = \frac{n}{N}$ Deci, „legea numerelor mari“ este o lege statistică obiectivă, constatîndu-se că anumite însușiri esențiale într-o colectivitate omogenă, se situează în bună parte în jurul mediei aritmetice ponderate. Sau, cu alte cuvinte: în populațiile mari unele fenomene se repetă aproximativ uniform; sau: există o oarecare stabilitate statistică în populațiile mari (atît timp cît nu intervine o schimbare a condițiilor); sau: $\bar{x}_i$ a unui număr suficient de mare de variabile aleatorii x_i își pierde caracterul aleator.

2. Se va avea în vedere logica determinărilor, în funcție de sex, vîrstă etc., de ex.:

— nu poate fi concludentă o cercetare a nașterilor după numărul femeilor din zonă, ci numai după numărul femeilor de vîrstă fertilă;

— nu poate fi concludentă cercetarea cancerului uterin după numărul femeilor din zonă, ci numai după un anumit interval de vîrstă;

— experimental s-a constatat că în 15 minute se pot reține de elevii dintr-o școală 5 idei noi, dar după 45 minute nu s-au putut reține 15 idei noi, din cauza oboselii.

Scanned with OKEN Scanner

3. Se va avea în vedere economicitatea cercetării. În acest scop la începutul cercetării se ia un microeșantion cu care se fac primele probe (măsurători, experiențe etc.) în vederea obținerii eventualelor deficiențe, după care se elimină sau se modifică unele analize, întrebări etc., stabilindu-se apoi precis datele de care este nevoie pentru cercetare. Institutele pentru sondarea opiniei publice din unele țări (ex. Institutul Gallup din S.U.A.) au o serie de întrebări bine chibzuite și experimentate cu care se adresează populației, inclusiv o serie de anchetatori profesioniști, cu experiență. Un astfel de exemplu este următorul: pentru a depista frecvența sifilisului dintr-o zonă geografică, numărul analizelor se poate reduce de 10 ori prin analizarea unui sînge „compus", recoltat de la 10 persoane într-o singură eprubetă, deoarece reacția este sensibilă și la diluția 1/10. În caz de pozitivitate, printr-o cercetare dichotomică se poate găsi cu ușurință cazul pozitiv. Situația este analogă cu analiza apelor reziduale dintr-un canal (care reprezintă un amestec de mai multe probe individuale), pentru a descoperi existența de excretori de germeni în zona respectivă.

4. Animalele de experiență se aleg în funcție de scopul urmărit; redăm cîteva exemple ce ni se par mai interesante:

— determinările de insecticide se fac pe muște femele, care sînt mai rezistente decît masculii;

— cercetătorii japonezi au demonstrat acum 60 de ani că gudronul este cancerigen, datorită faptului că au făcut experiențele pe iepuri și șoareci. Dacă ar fi utilizat cobai și șobolani ar fi conchis că gudronul nu este cancerigen, aceste animale fiind rezistente;

— coloranții azoici sînt cancerigeni pentru ficatul șobolanilor, dar sînt practic inactivi pentru iepure;

— alfa-naftil-amina este cancerigenă pentru vezică, avînd o perioadă de latență de 1 an pentru șobolan, de 2 ani pentru iepure, de 3—5 ani pentru cîine și uneori de 25 de ani pentru om;

— primele cercetări asupra talidomidei (contergan) au fost făcute pe șoareci cu rezultate bune. Ca urmare, această substanță produsă în 1959 s-a fabricat în 1961 sub diferite denumiri în diferite țări, utilizîndu-se drept calmant în vărsăturile incoercibile ale sarcinei. Dar, în no-

Scanned with OKEN Scanner

iembrie 1961 Dr. Lenz, pediatru din Hamburg, a sesizat statistic legătura dintre cauză și efect, ca urmare talidomida a fost scoasă din uz la 8 zile de la această sesizare.

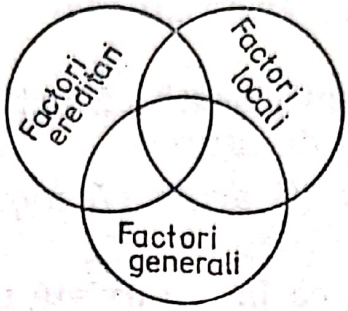


FIG. 22.

Pînă atunci, talidomida a produs 3 000 decese ale nou-născuților, 2 000 de invalidități grave în R.F.G., 1 200 în Anglia, 700 în Japonia etc., în total fiind aproximativ 6 000 de copii cu afecțiuni teratogene. Ulterior s-a constatat că șoarecii și iepurii nu sînt sensibili, iar șobolanul și omul sînt sensibili la talidomidă. Mai menționăm aici: cobaiul este sensibil la scorbut; iepurele este foarte sensibil la arterioscleroză; șobolanul este destul de rezistent la infecții; șoarecele nu are pigment melanic în piele.

5. Alegerea testelor trebuie să fie judicioasă. Testele trebuie să fie semnificative, eventual să se intersecteze, de ex. caria dentară este consecința mai multor factori cu-

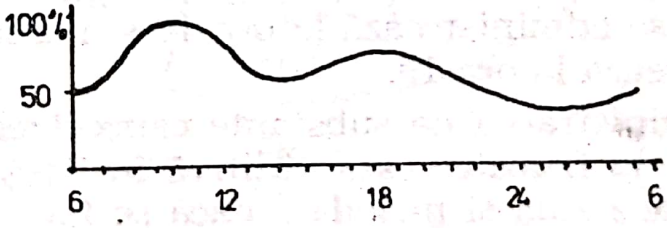


FIG. 23.

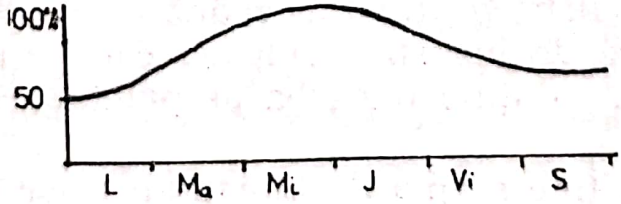


FIG. 24.

noscuți (pînă acum), coexistenți: ereditari, locali (igiena bucală) și generali (apă și alimente). Reprezentarea grafică este redată în fig. 22.

6. Se va ține seama și de bioritmul populației cercetate; cităm cîteva exemple ce ni se par mai interesante:

— Performanțele fizice și psihice ale omului sînt mari la mijlocul săptămînii (fig. 23), dar variază și în decursul zilei (fig. 24).

— Unele observații arată că activitatea cea mai rodnică este în a 3-a oră de muncă.

— Unele observații arată că memoria este mai bună între orele 14 și 16.

— S-a constatat că la om există un ciclu de 23 de zile (fizic), unul de 28 de zile (emoțional) și unul de 33 de zile (intelectual), deci există un ciclul de 28±3 zile (pe această bază se calculează bioritmul individual).

— Datorită semnalelor cronobiologice trimise de hipofiză diferitelor glande, există un ciclu de 21 de zile (tiroidian), un ciclu de 26 de zile (suprarenalian) și un ciclu de 28 de zile la femeie (catamenial), care dispare în cazul unor stresuri puternice, ex. prizonierat, închisoare etc.

— Efectul cel mai puternic la ingerarea alcoolului este între 6 și 10 dimineața, dar senzația de beție este dată mai ales la ingerarea alcoolului seara tîrziu.

— Lidocaina produce o anestezie dentară de 7—9 minute între orele 7 și 9 și mai mult de 30 minute după orele 15.

— Efectul antihistaminic al ciclopropandienei durează 16 ore dacă se administrează la ora 16 și numai 7 ore dacă se administrează la ora 19.

— Administrarea de substanțe cancerigene în glanda submaxilară la hamster este pozitivă în 72% cazuri dacă injecția se face ziua și niciodată dacă se face noaptea.

— Ingestia de glucide dimineața duce la o metabolizare (și la o curbă grafică) asemănătoare omului sănătos, spre deosebire de ingestia aceleiași cantități de glucide la amiază cînd se obține o curbă asemănătoare cu a unui diabetic.

— Se va ține seama de bioritmul unor organe, a căror activitate variază în decursul celor 24 de ore (fig. 25). De ex. activitatea cea mai intensă a ficatului este între orele 1 și 3, activitatea cea mai scăzută a plămînului este orele 4 și 5, calcemia cea mai scăzută este între orele 10 și 12, hemoglobina cea mai multă este între orele 11 și 13,

Scanned with OKEN Scanner

accidentele cardiace sînt mai frecvente între orele 19 și 21 etc. Aceste date corespund într-o oarecare măsură cu datele luate în considerare de acupunctura chinezească, ce consideră că activitatea unor organe este mărită în funcție

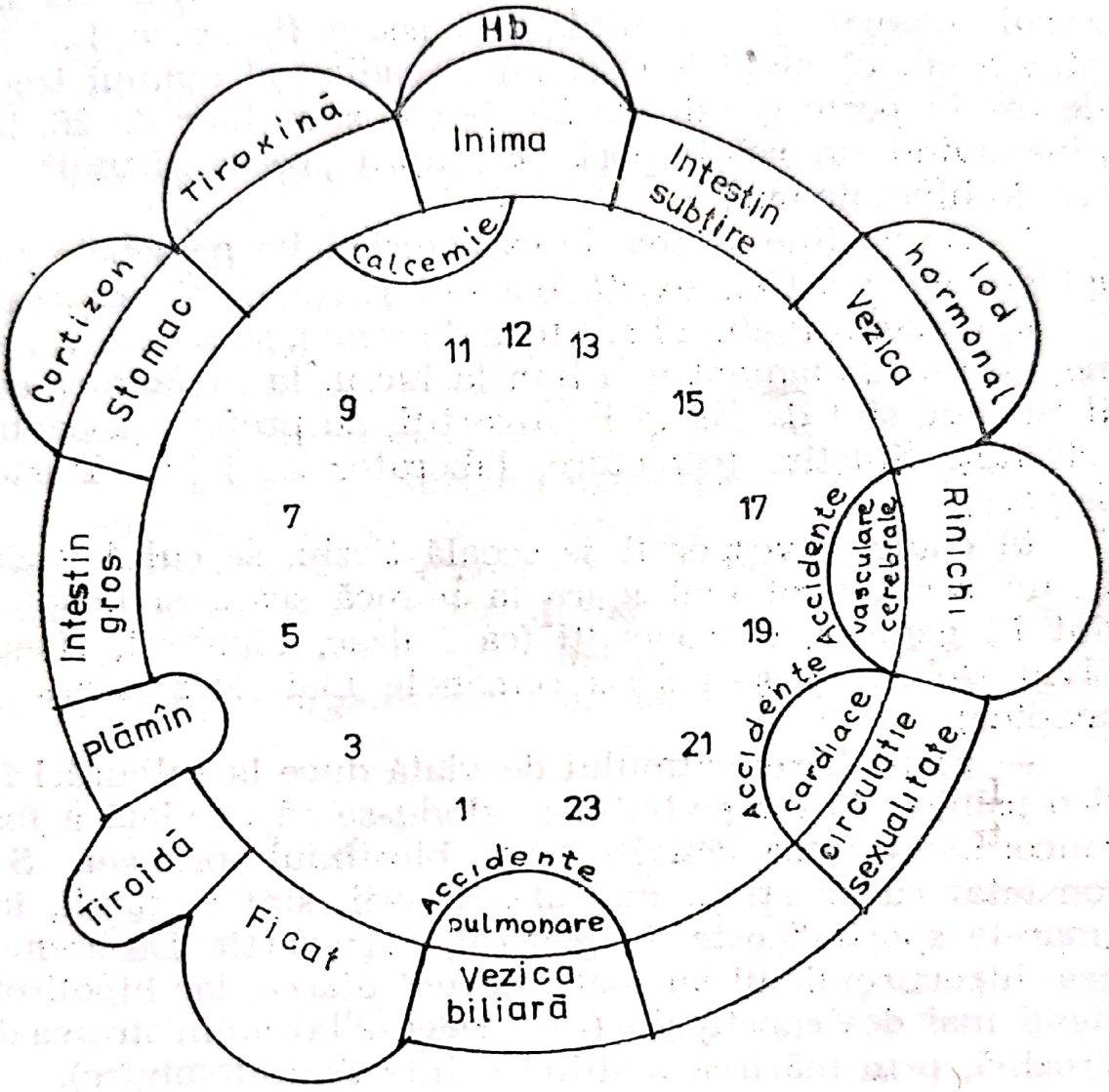


FIG. 25.

de orele zilei. S-a mai constatat că leucocitele circulante sînt maxime în primul sfert al zilei și minime noaptea, neutrofilele scad la ora 20 (fiind crescute la ora 13,30), monocitele și eozinofilele scad la ora 13,30 (fiind crescute la ora 02). De aici rezultă că recoltarea probelor de laborator ce se fac „dimineața pe nemîncate", ca și „3 pilule pe zi" trebuie acordate cu bioritmul omului, iar tratamentul să fie aplicat conform cronofarmacologiei pentru a avea efect maxim.

— Schimbările bruște de fus orar, mai ales la călătoria cu avionul, duc la oboseală, accese paradoxale de somn,

depresiuni mintale, tulburări de dispoziție etc., resincronizarea bioritmului necesitînd 2—3 săptămîni. Lucrul de noapte în industria modernă este adesea greu suportat, la orele mici vigilența slăbește. Benodiazepina injectată la animale de experiență dimineața (la orele 6—9) avansează orarul biologic, iar administrată seara (la orele 12—24) întîrzie orarul biologic. Ritmul circadian al omului izolat (de ex. în peșteri) este de 25 de ore sau chiar de 26, lumina avînd un rol important asupra „ceasornicului" situat în hipotalamus.

— Se va ține seama de caracteristicile personale psihofiziologice; astfel, există oameni matinali și vesperali:

a) oamenii matinali se culcă devreme, se scoală devreme, au un randament mai bun la lucru, la învățătură etc. dimineața, sînt de obicei introvertiți, au preferință pentru activități liniștite (cercetare, laborator etc.) ca Tolstoi, Selye.

b) oamenii vesperali se scoală tîrziu, se culcă tîrziu, au un randament mai mare la muncă seara și noaptea, sînt în general, extrovertiți (ca Balzac, Flaubert, Hegel. Liviu Rebreanu și-a scris romanele mai ales seara și noaptea).

— Schimbarea ritmului de viață duce la tulburări fizico-psihice, la longevivi observîndu-se că în viață a fost concordanță între activitate și bioritmul propriu. S-a constatat că sportivii, mai ales tinerii, sînt sociabili, iar „marele sportiv" este, în general, extrovertit. De asemenea, hipertiroidienii au mai frecvent diaree, iar hipotiroidienii mai des constipație (care cedează la administrarea de tirozină, prin mărirea motilității intestinului subțire).

— Se va ține seama de bioritmul animalelor de experiență; la animalele ce hibernează se produc schimbări profunde de viață: la marmotă ritmul cardiac de la 70/minut scade la 2—4/minut în hibernare, la arici de la 40—50/minut la 6—8/minut în hibernare etc.

— Se va ține seama de anotimp: experiențele de dozare a digitalei pe broască arată că variază cantitățile relative de la 1 310 vara la 3 300 unități iarna.

— Diferența de temperatură între iarnă și vară are de asemenea importanță: cînd aceasta este de aproximativ 11,5°C plajele sînt invadate de meduze, la Villefranche; cînd diferența este de peste 13°C sînt meduze puține. S-a observat că anii cu meduze multe sînt precedați de ani

Scanned with OKEN Scanner

secetoşi (la Nisa), existînd perioade de 4—5 ani cu meduze multe, urmate de 4—5 ani cu meduze puţine.

— Din p.d.v. a fotosensibilităţii, oamenii pot fi grupaţi astfel:

a) capacitate de pigmentare slabă sau nulă (categorie din care fac parte persoanele ce fac mai uşor cancer cutanat);

b) capacitate de pigmentare normală;

c) pigmentaţi natural (negri).

— În funcţie de unele date bioritmice, s-a ajuns a se prevedea cu o oarecare certitudine statistică sexul fătului.

— La investigarea unor funcţii fiziologice trebuie ţinut seama şi de unii factori naturali. Erupţiile (petele) solare care au o ritmicitate de 11,1 ani produc diverse efecte ca: aurore polare, perturbaţii magnetice, erupţii vulcanice, cutremure de pămînt, mai accentuate în perioadele de minime solare (fig. 26). Se consideră că ionizarea atmosferică (manifestată la scară mare şi vizibilă sub forma aurorelor polare), clima (uscată sau umedă) au cea mai mare influenţă asupra organismului uman. Persoanele cu meteo-

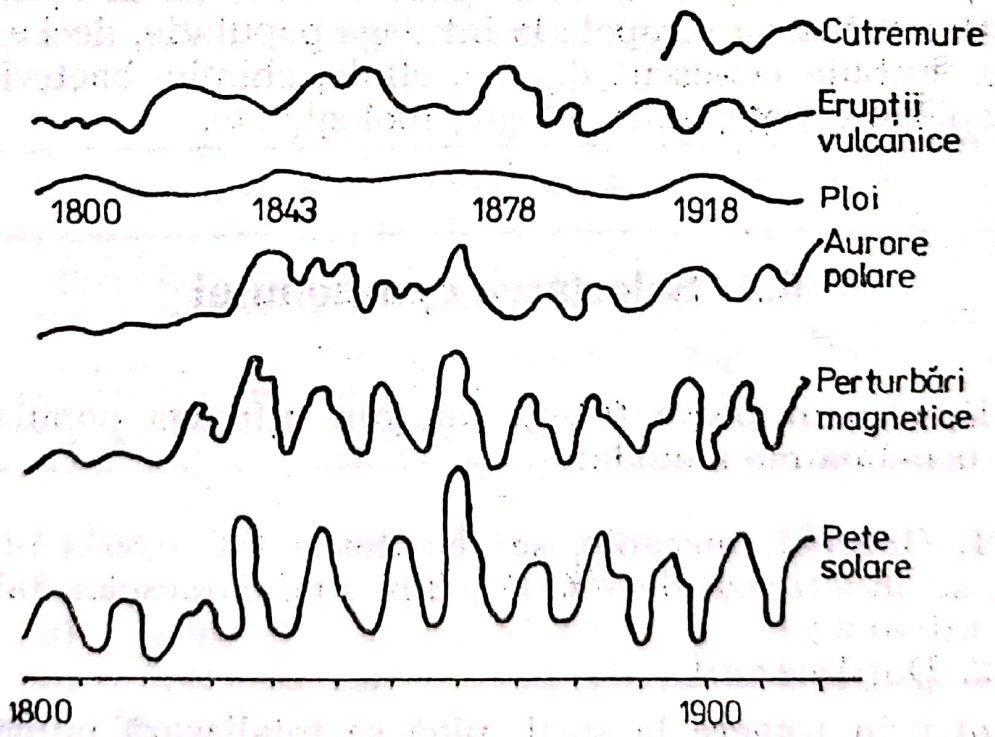


FIG. 26.

rosensibilitatea cea mai accentuată sînt bolnavii reumatici (apariţia reumatismului este mai frecventă primăvara), ulceroşii (cu agravare primăvara şi toamna şi hemoragii în februarie), cardiovascularii (lunile cele mai defavora-

bile fiind februarie și martie, iar decesele cele mai frecvente în iulie și august). Anotimpurile cu temperatura ridicată sînt nefavorabile pentru epileptici și histerici, vremea uscată este nefavorabilă parkinsonienilor, vremea umedă și rece este nefavorabilă reumaticilor și diabeticilor, presiunea atmosferică crescută este nefavorabilă hipertiroidienilor. Flebitele postoperatorii sînt mai frecvente primăvara și toamna. Femeile sînt cu 20% mai sensibile la radiațiile ultraviolete decît bărbații, dar vara sensibilitatea la aceste radiații este mai scăzută decît iarna și primăvara; foarte sensibili la aceste radiații sînt eczematoșii, hipertensivii și bolnavii de Basedow. O observație personală arată că uneori meteorosensibilitatea este ereditară: bunica, mama și fiul sînt sensibili și prevăd apropierea ploii. Tulburările meteorotrope se manifestă prin dureri, oboseală, insomnie, incapacitate de muncă, anxietate, amețeală, greață, vărsături etc.; deci cercetările statistice trebuie să țină seama de aceste fenomene pentru a ajunge la concluzii adevărate.

Exprimîndu-se în termeni cibernetici, este importantă entropia informațională ce se poate obține de la eșantion pentru a o putea extrapola la întreaga populație, deci eșantionul trebuie cunoscut d.p.d.v. clinic, chimic, bacteriologic, epidemiologic, antropologic, zoologic etc.

6.2. Selectarea eșantionului

Eșantionul poate fi selectat (din întreaga populație) prin următoarele metode:

1. Ordonat, mecanic, se recoltează tot a 50-a piesă, pînă se totalizează de ex. 1% din totalul pieselor fabricate într-o zi;

2. După hazard:

a) prin tragere la sorți, pînă se totalizează numărul necesar;

b) prin extragere oarbă (după STAS 6085-72), anume extrăgînd la întîmplare elemente din partea de sus, din mijloc și din partea de jos a stivei, rafturilor etc., pînă se totalizează numărul necesar;

Scanned with OKEN Scanner

c) prin metoda sferturilor (după STAS 1850-62), cînd eșantionul se formează printr-un șir de operații de amestecare și reducere, materialul împărțindu-se în sferturi, din care în sfîrșit se recoltează un sfert;

d) după un tabel cu numere aleatoare (STAS 6085-72), care se utilizează astfel: se înscriu în ordine (1, 2, 3, 4 . . .) numele persoanelor din zona de studiat, apoi se examinează persoana cu nr. 8735, apoi nr. 2552 etc., pînă se totalizează numărul necesar de persoane. Înainte vreme se utiliza „metoda Monte Carlo", care consta în utilizarea numerelor întîmplătoare apărute la ruleta din Monte Carlo. În lipsa unui astfel de tabel, ne putem confecționa unul, astfel: se ia de ex. numărul 4567 care se ridică la patrat=20 8574 89; de aici se iau cele 4 cifre din mijloc și se ridică la patrat: 8574^2=75 5134 76 și operația se repetă, obținîndu-se o serie de cifre: 4567, 8574, 5134, 5379 etc. care se întrebuințează ca mai sus. Datele respective se tabelează, în vederea prelucrării, într-un tabel ce trebuie să aibă:

— titlu, simplu, dar concis, asemănător cu subiectul dintr-o propoziție;

— coloane, care reprezintă stadiile, etapele, modificările, ce le are subiectul, asemănător cu predicatul dintr-o propoziție, de exemplu:

MALARIA, EVOLUȚIA ÎNTRE ANII 1950—1962

Coloane	1	2	3	4	5
Ani	1950	1955	1960	1961	1962
Cazuri noi	55 175	308	78	28	—

Coloanele se numerotează pentru referiri ulterioare.

În rubricile respective, notarea se face astfel:

a) se scrie cifra respectivă cînd avem date;

b) se trece semnul — cînd rezultatul este negativ sau lipsește fenomenul;

c) se scriu puncte . . . cînd lipsesc date;

d) cînd tabelul este citat dintr-o lucrare, se indică sursa bibliografică (de ex.: Breviar de statistică medico-sanitară, 1967).

Scanned with OKEN Scanner

6.3. Stabilirea volumului eșantionului

1. În general, eșantionul reprezintă 1 . . . 5% din totalul populației. Dovadă că eșantionul a fost bine ales este faptul că dacă se cercetează 1 . . . 2 . . . 3% din populație, nu apar diferențe mari la rezultatele obținute.

2. În cazul unui control al calității unor produse industriale, se recoltează 5—10% din piesele fabricate și doar 2% în cazul unei serii de fabricație foarte mari.

3. Volumul eșantionului se stabilește în funcție de nivelul de încredere ce urmează să-i acordăm. Exemplu: pentru a avea un interval de încredere de 95% și precizia pînă la 0,3 s (abateri patratice), trebuie să avem 46 elemente (cazuri, experiențe etc.). Dacă ne mulțumim cu o abatere de ± 1 s vom citi în tabelul nr. 12 că ne sînt suficiente 7 elemente.

4. Pe baza regulii „celor 3 sigme" s-a întocmit tabelul nr. 13, în care este redată legătura dintre nivelul de încredere și numărul n de elemente din eșantion. Astfel, pentru un nivel de încredere de p=0,985 trebuie să avem minim 10 elemente.

5. Plecînd de la formula $ES=\frac{s}{\sqrt{n}}$ numărul elementelor se poate stabili în funcție de abaterea ce o tolerăm, după formula: $n=\frac{s^2}{(ES)^2}$.

Exemplul 1:

Ce număr n de elemente trebuie să avem în eșantion dacă admitem că $s=\pm 10$ și ES=2?

$$n=\frac{10^2}{2^2}=\frac{100}{4}=25 \text{ elemente}$$

Exemplul 2:

Ce număr n de elemente trebuie să avem în eșantion dacă admitem că $s=\pm 10$ și ES=1?

$$n=\frac{10^2}{1^2}=\frac{100}{1}=100$$

Interpretarea se face astfel:

Scanned with OKEN Scanner

Cu cît sîntem mai severi, cu atît admitem un ES mai mic, caz în care sînt luate în considerare numai 66% din elemente (ES=1), celelalte care sînt „mai departe" de $\bar{x}$ vor fi neglijate. Deci vom fi mai toleranți cînd știm că nu putem obține multe elemente în eșantion (elementele se distrug, sînt puțini bolnavi în spital etc.) și admitem ± 2 sau 3 ES. Totuși, dacă numărul elementelor este prea mic, riscăm să avem rezultate eronate; de ex. dacă din 2 elemente pe care le avem la dispoziție, unul nu este corespunzător, nu înseamnă (prin extrapolare) că 50% din populație este necorespunzătoare.

6. Pentru atingerea unui interval de încredere, se poate estima numărul necesar de măsurători, dacă se adoptă apriori o valoare pentru raportul $q=\frac{A_a}{s}$, unde:

A=abaterea absolută față de $\bar{x}$
s =abaterea eșantionului

Exemplu:

Diametrul mediu $\bar{x}$ a unor piese=15 mm$\pm$3 mm.

Ne impunem ca A_a să fie $\pm$1 mm (deci să avem 15$\pm$1 mm).

Cîte elemente trebuie să avem în eșantion?

$$q=\frac{A_a}{s}=\frac{1}{3}=0{,}3.$$

Din tabelul nr. 12 rezultă că la q=0,3 și p=90 sînt necesare 32 de elemente în eșantion.

7. Numărul de elemente din eșantion se mai poate afla plecînd de la formula: $k=\sqrt{\frac{n}{Q}}\cdot 100$, de unde: $n=Q\cdot\left(\frac{k}{100}\right)^2$, unde:

n =numărul elementelor din eșantion
Q=numărul de piese executate între două recoltări
k =numărul de piese controlate

Exemplu:

Într-o oră s-au fabricat 500 piese, din care s-au controlat 50=10%. Să se stabilească numărul n al eșantionului.

$$n=500\cdot\left(\frac{10}{100}\right)^2=500\cdot 0{,}01=5 \text{ piese.}$$

Capitolul 7

Probleme rezolvate

7.1. Obținerea de informații dintr-o singură valoare

Exemplul 1:

Fie valoarea: 10. Să se prelucreze statistic.

Răspuns: Nu se poate obține altă informație în plus, deoarece nu se știe ce reprezintă, de unde provine, ce semnificație are, ce legături conexe are cu alte valori etc.

Exemplul 2:

Dintr-o stivă de saci s-a cîntărit un sac, care a avut 50 kg. Să se prelucreze statistic pentru a stabili greutatea celorlalți saci.

Răspuns: Nu se poate obține o altă informație în plus. Conform diferitelor metode statistice, avem nevoie de un număr minim de elemente dintr-o mulțime, deci de un eșantion, pentru a putea caracteriza întreaga mulțime.

7.2. Obținerea de informații din două valori

Exemplul 1:

Fie valorile: 10; 20. Să se prelucreze statistic.

Răspuns:

a) raportul dintre ele este: $10 : 20 = 0{,}5$ sau $20 : 10 = 2$;

b) raportul procentual este: 50% sau 200%;

c) $\bar{x} = (10 + 20) : 2 = 15$;

d) s nu se poate stabili, deoarece sînt numai 2 valori.

Scanned with OKEN Scanner

7.3. Şir progresiv geometric

Exemplul 1:

O „capcană" matematică este următoarea problemă: un nufăr îşi dublează suprafaţa frunzelor aşa încît în 10 zile ocupă toată suprafaţa lacului. Întrebarea este: în a cîta zi va fi ocupată 1/2 din suprafaţa lacului?

Deşi sîntem tentaţi să răspundem — în a cincea zi —, adevărul îl stabilim după întocmirea următorului tabel:

ZIUA	SUPRAFAŢA
1	1
2	2
3	4
4	8
5	16
6	32
7	64
8	128
9	256
10	512 = 100%

Deci 1/2 (50%) este ocupată în ziua a noua.

Exemplul 2:

O variantă a acestei probleme este următoarea:

Un nufăr îşi dublează suprafaţa frunzelor încît în 10 zile ocupă suprafaţa unui lac. Întrebarea este: în cîte zile va fi ocupată suprafaţa lacului, dacă sînt 2 nuferi? De asemenea sîntem tentaţi să răspundem: în $\frac{1}{2}$ din timp. Dar, după tabel:

ZIUA	NUFĂR 1	NUFĂR 2	NUFĂR 1+2
1	1	1	2
2	2	2	4
3	4	4	8

Scanned with OKEN Scanner

ZIUA	NUFĂR 1	NUFĂR 2	NUFĂR 1+2
4	8	8	16
5	16	16	32
6	32	32	64
7	64	64	128
8	128	128	256
9	256	256	512
10	512	512	1 024=100%

rezultă că 1/2 din suprafață va fi ocupată în ziua a noua.

Exemplul 3:

O foaie de hîrtie are grosimea de 0,1 mm. Dacă prin împăturire de 5 ori obținem 32 de foi, grosimea pachetului va fi 32·01=3,2 mm.

Întrebarea este: dacă vom împături foaia de hîrtie de 20 de ori, ce grosime va avea pachetul?

Problema se rezolvă astfel:

dacă $n_5=2^5=32\cdot 0{,}1=3{,}2$ mm, atunci
$n_{20}=2^{20}$ și prin logaritmare avem:

$\log n_{20}=20\cdot\log 2=20\cdot 0{,}3=6{,}0$. Antilog 0=1, deci vom avea 7 cifre întregi (caracteristica fiind 6), așadar rezultatul este: $n_{20}=1\,000\,000$ (foi)$\cdot 0{,}1$ mm=100 m.

7.4. Medii

Exemplul 1:

La ducere într-o localitate, viteza unui automobil a fost 60 km/h, iar la întoarcere a fost 40 km/h. Care a fost viteza medie?

Sîntem tentați a răspunde: 50 km/h=media aritmetică simplă. Dar, de fapt trebuie să calculăm media armonică, astfel:

$$\bar{x}_{arm\,s}=\frac{n}{\frac{1}{x_1}+\frac{1}{x_2}}=\frac{2}{\frac{1}{60}+\frac{1}{40}}=\frac{2}{0{,}0416}=48 \text{ km/h.}$$

Demonstrația se poate face în felul acesta: dacă 120 km au fost parcurși cu 60 km/h (la ducere) și 120 km au fost parcurși cu 40 km/h (la înapoiere), avem:

120 : 60=2 ore
120 : 40=3 ore

5 ore, astfel că 120 km : 5 ore=48 km/h.

Exemplul 2:

Un produs, de ex. ardeiul a fost :

2	buc./leu	primăvara
3	„	vara
10	„	toamna

Care este media prețurilor?

Calculînd, obținem: (2+3+10) : 3=5 bucăți/leu, deci 1 leu : 5 buc.=0,2 lei.

Dar făcînd calculul pe sortimente avem:

$$\frac{1 \text{ leu}}{2 \text{ buc.}} = 0{,}50 \text{ lei/bucată}$$

$$\frac{1 \text{ leu}}{3 \text{ buc.}} = 0{,}33 \text{ lei/bucată}$$

$$\frac{1 \text{ leu}}{10 \text{ buc.}} = 0{,}10 \text{ lei/bucată}$$

0,93 lei : 3=0,31 lei/bucată.

Deci rezultatele (0,20 lei și 0,31 lei) nu sînt egale. Deoarece prețurile sînt invers proporționale cu numărul bucăților, trebuie utilizată media armonică simplă, astfel:

$$\bar{x}_{arm\ s} = \frac{3}{\frac{1}{2} + \frac{1}{3} + \frac{1}{10}} = \frac{3}{0{,}93} = 3{,}2 \text{ bucăți/leu.}$$

Dacă s-ar fi știut și numărul total al bucăților cumpărate, s-ar fi utilizat media armonică ponderată, caz în care prețul/bucată ar fi fost mai mic (decît 3,2 bucăți/leu), deoarece ponderea ar fi avut-o cantitatea cea mai mare (aici 10 buc./leu).

În statistică se mai întrebuințează și valoarea „mediana", care este termenul din mijlocul seriei, după ce

termenii au fost aranjați (ca valoare) în ordine (crescîndă sau descrescîndă). Este asemănătoare cu media aritmetică simplă. De asemenea mai există termenul de „modúl" sau „dominantă", care este valoarea cu frecvența cea mai mare. Este asemănătoare cu $\bar{x}_{ap}$.

7.5. Abaterea patratică

Exemplul 1:

Fie valoarea: 10; 12. Să se determine $\bar{x} \pm s$.

$$\bar{x} = (10+12) : 2 = 11$$

s nu se poate calcula decît dacă avem cel puțin 3 valori.

Exemplul 2:

Fie valorile: 1; 10; 100. Să se determine $\bar{x} \pm s$

$$\bar{x} = (1+10+100) : 3 = 111 : 3 = 37$$

Rezultatul nu este corect, căci valorile sînt în progresie geometrică, deci trebuie calculată media geometrică:

$$\sqrt[3]{1 \cdot 10 \cdot 100} = \sqrt[3]{1\,000} = 10$$

s nu se determină decît la media aritmetică, dacă valorile provin dintr-o distribuție gaussiană.

Exemplul 3:

Fie valorile: 1; 1; 1; 3; 50. Să se determine $\bar{x} \pm s$. Calculîndu-se s-a găsit: $\bar{x} \pm s = 11{,}2 \pm 21$.

Comentarii:

— s mare se datorește valorii 50, deci se pune problema eliminării;

— valoarea aberantă se îndepărtează numai dintr-o distribuție normală, deci:

— distribuția nefiind normală, valoarea 50 nu trebuie eliminată;

— s se calculează numai la distribuțiile normale.

Scanned with OKEN Scanner

7.6. Testul t

La începutul acestui secol, statisticianul W. Gosset a fost angajat la o fabrică de bere unde trebuia să compare diferitele calități de bere. În 1908 a ajuns la o formulă pe care a publicat-o sub pseudonimul „Student".

John Lawes (1804—1900) a înființat în 1843 ferma agricolă „Rothamstedt Experimental Station", pe care a donat-o mai tîrziu statului împreună cu 100 000 lire sterline, pentru asigurarea cercetărilor (asolamente, îngrășăminte, ameliorarea raselor etc.). La această stațiune în anul 1919 a fost angajat statisticianul R. A. Fisher, pentru a prelucra enorma cantitate de date adunate pînă atunci în arhivele stațiunii. Rodul muncii lui s-a concretizat în 1925, cînd a publicat „Metode statistice pentru cercetătorii științifici", iar în 1935 „Proiectarea experimentelor". Începînd din 1922, la această stațiune a lucrat și Gosset; din colaborarea lor a rezultat „testul t Student și Fisher".

Exemplul 1:

Într-o populație cu n=200 indivizi, cercetătorul A a găsit $\bar{\bar{x}} \pm \sigma = 20 \pm 5$. Cercetătorul B a dispus numai de un eșantion cu n=25 indivizi și a găsit $\bar{x} \pm s = 19 \pm 4$. Întrebarea este: putem estima ce valori a găsit cercetătorul A (presupunînd că nu cunoaștem rezultatele sale)?

Problema se rezolvă astfel:

$$\bar{\bar{x}}_t = \bar{x} \pm (ES \cdot t) = 19 \pm \left(\frac{4}{\sqrt{25}} \cdot t\right) = 19 \pm (0{,}8 \cdot t).$$

Din tabelul nr. 2 aflăm că la 25—1=24 gr lib și p=0,05, t=2.

Deci avem: $\bar{\bar{x}}_t = 19 \pm (0{,}8 \cdot 2) = 19 \pm 1{,}6$ la p=0,05.

În cazul dat, valorile 17,4 . . . 20,6 se găsesc în plaja $\bar{\bar{x}} \pm \sigma = 20 \pm 5$ (adică în zona 15 . . . 25) în 95% din cazuri.

7.7. Testul χ^2

Johann Mendel s-a născut la 22 iulie 1822 la Heizendorf (Moravia de nord), fiind al doilea fiu al unui fermier sărac. După terminarea liceului, în 1843, la 21 de ani se

călugărește la mănăstirea din Brünn (azi Brno — Cehoslovacia), luînd numele de Gregor. La 1850 se prezintă la un examen pentru a deveni institutor, dar nu reușește. Starețul mănăstirii îl ajută să ocupe un post de profesor suplinitor la liceul din Znojmo. În 1851 ordinul călugăresc îl trimite la Universitatea din Viena, unde studiază matematica, fizica și științele naturii. Reîntors la Brno, rămîne suplinitor, deoarece nu reușește la examenul de definitivat. Începînd cu anul 1856 s-a ocupat cu experiențe de încrucișări de mazăre și la 8 februarie prezintă unica sa comunicare: Experiențe asupra hibrizilor la plante. Comunicarea a fost considerată „interesantă", dar fără aplicații practice, așa că în procesul verbal s-a consemnat: nu au fost întrebări, nici discuții, lucrarea fiind „pusă la dosar". Aceasta s-a datorat în bună parte și faptului că rezultatele nu au fost prelucrate statistic pentru a vedea importanța lor. Din această cauză lucrarea a rămas ignorată timp de 35 de ani, poate și pentru faptul că a apărut „prea devreme pentru timpul său". După alegerea sa ca stareț în 1868, aceste cercetări au fost reduse, apoi abandonate. Mendel a murit la 6 ianuarie 1884 în urma unei boli de inimă și rinichi de care suferea de mulți ani. În fața fostei sale chilii, azi muzeu memorial, se cultivă permanent mazăre, în semn de omagiu. În 1881, botanistul Focke citează lucrarea lui Mendel într-un volum, de care iau cunoștință olandezul Hugo de Vries, germanul Correns și austriacul Tschermak. Karl Pearson introduce testul χ^2 în 1900, cînd Batesson traduce lucrarea lui Mendel în engleză, denumind știința respectiv, „Genetică". Azi lucrarea lui Mendel are aplicație în genetica plantelor, genetica animalelor, genetica populațiilor, biologia moleculară, biotehnologie etc.

Într-una din cercetările sale pe 1 074 plante, Mendel a observat că 787 erau înalte și 287 erau scunde, deși el aștepta 805,5 (75%) înalte și 268,5 (=25%) scunde, dar s-a mulțumit să constate. Să restabilim adevărul utilizînd testul χ^2.

ŞIR	PLANTE	OBSERV	AŞTEPT	(OBS.-AŞT.)	(OBS.-AŞT)²
	COL 1	2	3	4	5
1	înalte	787	805,5	—18,5	342,25
2	scunde	287	268,5	+18,5	342,25
Total		1 074	1 074		

$$\chi^2=\Sigma\frac{(F\ obs-F\ aşt)}{F\ aşt}=\frac{342,25}{805,25}+\frac{342,25}{268,25}=0,42+1,27=1,69$$

$$\text{Gr. lib.}=(col-1)\cdot(şir-1)=(2-1)\cdot(2-1)=1.$$

Cifra 1,69 se află la 1 gr. lib. la p=0,2, ceea ce înseamnă că 1—0,2=0,8=80% concordanţă şi doar 20% neconcordanţă.

Rezultatul se exprimă: χ^2: se acceptă ipoteza nulă, deoarece diferenţele dintre eşantioane sînt (aproape) nule. Cu alte cuvinte, Mendel nu a greşit cînd a aşteptat proporţia de 75%+25%, rezultatul observat fiind aproape egal cu cel dorit teoretic. Deci Mendel a apreciat bine mersul experienţei, dar neavînd argumente obiective, statistice, experienţele sale au fost considerate interesante teoretic şi . . . atît.

Exemplul 2:

Cercetătorul A are în studiu o analiză unde:

FO	FT
14	17
15	17
20	17

iar cercetătorul B are o analiză din acelaşi material cu rezultatele:

FO	FT
30	20
35	25
40	25

Întrebarea este: care rezultat va fi mai demn de încredere?

Răspuns: nu se poate face nici o comparație, deoarece FT sînt diferite (FO pot fi diferite).

Exemplul 3:

În eșantionul A am găsit: 7; 4; 9; 15, iar în eșantionul B am găsit: 8; 7; 15; 4. Întrebarea este: putem ordona cifrele, pentru o mai ușoară calculare, în ordine crescîndă (4; 7; 9; 15 și 4; 7; 8; 15)?

Răspunsul este: nu putem schimba ordinea cifrelor, deoarece trebuie să comparăm pe 7 cu 8, pe 4 cu 7 etc., dacă aplicăm testul χ^2. Ordinea cifrelor se poate modifica, dacă stabilim $\bar{x}$-urile și s-urile.

Exemplul 4:

La o cercetare s-au efectuat analize din aceeași substanță, dar cu metode diferite și s-a găsit:

FO_a	FO_b
5	6
7	8
10	12
9	5
31	31

Întrebarea este: există concordanță între eșantioane, adică între metode? Deoarece ambele eșantioane sînt „observate", fără a avea o frecvență teoretică, se procedează astfel:

$$\chi^2=\Sigma\frac{(F\,a-F\,b)^2}{\left(\frac{F\,a+F\,b}{2}\right)}=\frac{(5-6)^2}{\left(\frac{5+6}{2}\right)}+\frac{(7-8)^2}{\left(\frac{7+8}{2}\right)}+\frac{(10-12)^2}{\left(\frac{10+12}{2}\right)}+$$

$$+\frac{(9-5)^2}{\left(\frac{9+5}{2}\right)}=\frac{1}{5,5}+\frac{1}{7,5}+\frac{4}{11}+\frac{16}{7}=$$

$$=0,18+0,13+0,36+2,28=2,95$$

Gr. lib.=(col—1)·(șir—1)=(2—1)·(4—1)=1·3=3.

Cifra 2,95 la 3 gr. lib. o găsim la p=0,4, ceea ce înseamnă că există în 60% concordanță și 40% discordanță. Se recomandă a se repeta cercetările pentru a se preciza concordanța (și discordanța).

Exemplul 5:

Executîndu-se aceleași analize prin două metode, s-a găsit:

FOa	FOb
15	20
12	18
10	15
—	16

Întrebarea este: care metodă este mai bună?

Răspunsul este: nu se poate spune care metodă este mai bună (prin testul χ^2) și nici nu se poate aplica testul χ^2, deoarece nu sînt egale frecvențele. Dacă se renunță la valoarea 16, se poate aplica testul χ^2.

7.8. Stabilirea gradelor de libertate la testul χ^2

Prin stabilirea gradelor de libertate înțelegem precizarea raportului dintre elemente dependente și cele independente, dintr-o serie de valori.

Exemplul 1:

Fie următoarele valori: 10+20+30=60.

Logica stabilirii gr. lib. este: termeni independenți — termeni dependenți. În cazul dat, suma fiind stabilită (60), din cei 3 termeni, numai 2 sînt independenți, căci al 3-lea depinde de ceilalți 2, suma rămînînd aceeași. De ex. 20 depinde de 10 și 30, pentru ca suma să fie 60.

Deci gr. lib.=term. indep.—1=3—1=2.

Exemplul 2:

Fie următorul tabel pentru un test χ^2:

ŞIRURI	COLOANE			
	A	B	C	TOTAL
1	1	2	3	6
2	4	5	6	15
3	7	8	9	24
4	10	11	12	33
total	22	26	30	78

Aparent, avem 4 şiruri·3 coloane=12 valori. De fapt, avem mai puţine şi anume (totalurile fiind stabile):

şirul

1 : 2	valori indep.	+1	val. dep.	(de ex. 2 depinde de 1; 3; 6);
2 : 2	,,	+1	,,	(de ex. 6 depinde de 4; 5; 15);
3 : 2	,,	+1	,,	(de ex. 7 depinde de 8; 9; 24);
4 : 0	,,	+3	,,	(oricare din cifre).
6		3		

Acelaşi lucru este şi în cazul coloanelor (unde de asemenea totalurile sînt stabile):

coloana

A : 3	valori independente	+1	valoare dependentă
B : 3	,,	+1	,,
C : 0	,,	+4	,,
6		6	

Aceasta conduce la formula: gr. lib.=(col.—1)·(şir—1), în cazul nostru: (3—1)·(4—1)=2·3=6.

Exemplul 3:

Dintr-o populaţie cu $\bar{\bar{x}} \pm \sigma = 100 \pm 10$, am obţinut un eşantion la care am găsit următoarele valori: 90; 100; 110. Întrebarea este: aceste valori sînt în concordanţă cu ale populaţiei întregi? Deci aici trebuie să comparăm o serie

de frecvențe observate cu o frecvență teoretică. Deoarece cunoaștem pe $\bar{x}$, luăm această valoare drept valoare teoretică și avem:

F.O.	F.T.
90	100
100	100
110	100

$$\chi^2=\Sigma\frac{(90-100)^2}{100}+\frac{(100+100)^2}{100}+\frac{(110+100)^2}{100}=\frac{100}{100}+$$

$$+\frac{0}{100}+\frac{100}{100}=2$$

Gr. lib.=(col.—1)·(șir—1)=(2—1)·(3—1)=1·2=2.

În cazul de față mai există o restricție, adică se mai scade 1 gr. lib. pentru următorul motiv: deoarece ambele coloane provin din aceeași populație, se presupune că și s-urile sînt egale.

Cifra 2 la 1 gr. lib. se află la p=0,2, ceea ce înseamnă că eșantioanele se deosebesc în 20% din cazuri (și se asemănă în 80%). Această mare asemănare se datorește faptului că am luat din populația respectivă un eșantion reprezentativ, deci care a avut toate „atributele" (calitățile și defectele) populației întregi.

În concluzie:

1. Gr. lib.=valori independente—restricții.

2. Cînd singura restricție este suma (atît la șiruri cît și la coloane), gr. lib.=(col.—1)·(șir—1). Din această cauză, la testul χ^2 trebuie să avem minim 2 șiruri și 2 coloane.

3. Cînd $s_1=s_2$ cunoscut aprioric sau condiționat aprioric, chiar dacă nu se cunoaște valoarea cifrică a lui s_1 și s_2, se mai pierde 1 gr. lib.

4. Oradea frecvențelor (cifrelor) în șiruri și coloane trebuie să fie conformă cu regulile de comparat, deci nu se poate schimba ordinea acestora; în caz contrar se obțin rezultate eronate.

7.9. Problemă soluţionată prin mai multe metode

Exemplul 1:

Fie următoarele valori, găsite la o analiză: 40; 10; 5; 20; 60; 30; 40.

Întrebarea este: ce informaţii putem obţine prin prelucrarea acestor valori?

a) Reprezentînd grafic valorile în ordinea de apariţie a lor, se obţine o reprezentare ondulatorie a acestora

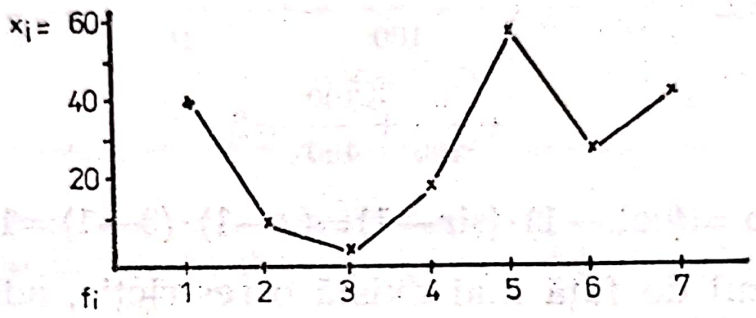


FIG. 27.

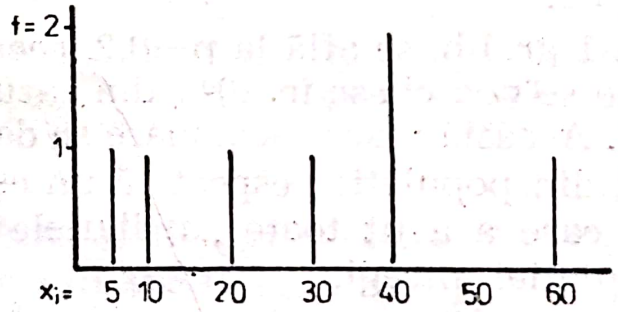


FIG. 28.

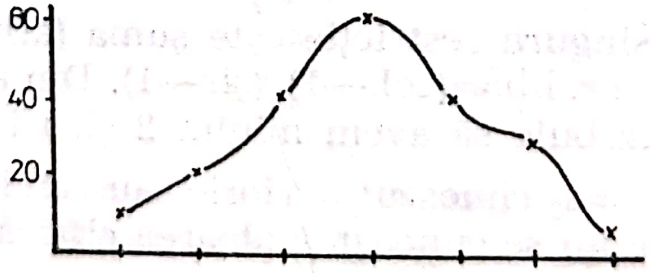


FIG. 29.

(fig. 27). Reprezentînd grafic frecvenţa de apariţie, se vede neregularitatea apariţiilor (fig. 28). Dar, ordonînd valorile în jurul valorii maxime, se obţine o curbă Gauss (fig. 29).

b) determinarea $\bar{x} \pm s$

n	x_i	$(x_i)^2$
1	5	25
2	10	100
3	20	400
4	30	900
5	40	1 600
6	40	1 600
7	60	3 600
7	205	8 325

$$\bar{x} = 205 : 7 = 29,3$$

$$s = \sqrt{\frac{8\,325 - \frac{205^2}{7}}{7-1}} = \sqrt{\frac{8\,325 - 6\,003}{6}} = \sqrt{\frac{2\,321}{6}} = \sqrt{386} = 19,6.$$

Rezultatul este: $\bar{x} \pm s = 29,3 \pm 19,6$.

Valoarea mare a lui s denotă o lipsă de omogenitate a populației. Se pune întrebarea dacă valoarea 5 nu trebuie îndepărtată din calcul, fiind prea mică. Verificînd cu ajutorul „regulei celor 3 sigme" constatăm: $\frac{x_i - \bar{x}}{s} = \frac{2 - 29,3}{19,2} = 1,2$. Deoarece $1,2 < 3$ (sigme), valoarea dubioasă $x_d = 5$ nu trebuie îndepărtată din serie.

c) Estimarea lui $\bar{\bar{x}}$.

Fiind valori puține, estimarea se face cu ajutorul testului t astfel: $\bar{\bar{x}} \pm (ES \cdot t) = 29,3 \pm \left(\frac{s}{\sqrt{n}} \cdot t\right) = 29,3 \pm \left(\frac{19,6}{\sqrt{7}} \cdot t\right) = 29,3 \pm (7,4 \cdot 2,4) = 29,3 \pm 17,4.$

Cifra 2,4 este valoarea lui t la $7 - 1 = 6$ gr. lib. și $p = 95\%$.

Interpretarea rezultatelor:

1. Înregistrarea cronologică a valorilor arată o apariție oscilantă, ondulatorie a acestora;

2. Reprezentarea grafică în funcție de frecvență arată o neregularitate în apariția valorilor;

3. Valorile sînt destul de disparate, dovadă s mare.

În concluzie:

1. Eșantionul nu pare să prezinte prea mare încredere, recomandîndu-se examinarea și a altor eșantioane (din aceeași populație statistică), de preferat a avea mai multe elemente;

2. Din curba lui Gauss (fig. 29) rezultă că este o populație „normală".

În încheiere putem spune că:

1. Din informații sărace, obținem răspunsuri sărace. Dacă avem o singură valoare, nu o putem prelucra statistic, deoarece nu o putem situa în context cu o valoare de referință. Deci, pentru a putea caracteriza un fenomen, trebuie să avem cel puțin două „atribute" sau caracteristici; de ex. temperatura bolnavului funcție de timp (=2 atribute), curentul electric funcție de tensiune și rezistență (=3 atribute), greutatea unei bare metalice funcție de lungime, grosime, densitate (=4 atribute) etc.

2. Trebuie știut raportul dintre caracteristici; de ex. $x=a \cdot b$; $x=a : b$; $x=a+b$ etc.

3. Înscrierea valorilor se face cronologic, ordine ce se menține atît în tabele, cît și în reprezentările grafice, ex. temperatura bolnavului, evoluția presiunii atmosferice etc. Alteori, valorile se ordonează în șir crescător sau descrescător, pentru:

— a face calculele mai ușor;

— a stabili dacă valorile extreme sînt aberante și a le elimina;

— a aprecia evoluția evenimentului, eventual stabilind media glisantă pentru a vedea tendința fenomenului (de creștere sau scădere).

Cînd avem elemente puține, concordanța dintre eșantioane se stabilește mai bine cu ajutorul testului t decît cu χ^2, corelația dintre aceste teste fiind următoarea: de ex.:

χ^2 la 1 gr. lib. și $p=0,1$ este 2,71

t la ∞ gr. lib. și $p=0,1$ este $1,64=\sqrt{2,71}$.

Alegere preferențială sau selectare opțională

În afara calculelor matematice, a căror rezultate sînt imparțiale, în viață apar situații cînd trebuie aplicată o

altă metodă de apreciere a situației, mai ales cînd avem a alege una din mai multe situații. Pentru tratamentul melanomului malign (o formă de cancer a pielii), au fost luate în considerare 36 de caracteristici: localizare, durată, extindere, adîncime etc., constatîndu-se că cel mai important factor este adîncimea (în grosimea pielei).

Stabilirea acestui lucru s-a făcut apelîndu-se la „analiza științifică a voturilor", stabilită în 1785 de matematicianul, omul politic și filosoful Condorcet, care a dat următorul exemplu:

23	alegători	votează	pentru	A, C, B
19	„	„	„	B, C, A
16	„	„	„	C, B, A
2	„	„	„	C, A, B
60				

S-ar părea că voturile cele mai multe le-a obținut A. Dar în realitate trebuie făcută comparația opțiunilor, comparînd doi cîte doi candidați, deoarece 23 alegători preferă pe A în primul rînd, apoi pe C și la urmă pe B etc. Tabelînd rezultatele, avem:

		AB	AC	BA	BC	CA	CB	TOTAL
23	ACB		23				23	46
19	BCA				19	19		38
16	CBA			16			16	32
2	CAB	2				2		4
60		2	23	16	19	21	39	120
		25		35		60		

Rezultatul este următorul:

A=25 : 2=12,5
B=35 : 2=17,5
C=60 : 2=30
60,0

Deci din 60 alegători, 30 au preferat pe C față de ceilalți candidați, deci C este favoritul.

Metoda poate fi utilizată la alegerea unei persoane din trei candidați (sau din mai mulți, cînd tabelul trebuie extins). Programînd pe un calculator, rezultatul apare în cîteva minute.

Scanned with OKEN Scanner

Încheiere

Terminarea calculelor și obținerea rezultatelor nu înseamnă că lucrul s-a încheiat. Datele respective trebuie interpretate, trăgîndu-se concluziile necesare atît pentru situația actuală, cît și pentru posibilele prognoze. Rezultatul statistic trebuie interpretat în context cu toți factorii care au concurat: fizici, chimici, demografici, epidemiologici, bacteriologici etc., așa cum un buletin de analiză se interpretează la patul bolnavului în funcție de temperatură, tratament, operație ș.a. La calculatorul electronic apare „Error" la un calcul matematic greșit (de ex. $\sqrt{-2} \longrightarrow$ Error), dar la o greșeală de logică nu apare „Error". În orice caz, un cercetător își sprijină concluziile asupra unei ipoteze statistice (=presupunere asupra unei situații statistice, ce poate fi adevărată) prin mai multe metode. Dar trebuie știut că testele de ipoteză duc la interpretări statistice și nu la interpretarea cauzelor. Deci trebuie luate toate măsurile, pentru ca rezultatele să aibă fiabilitate (=încredere în rezultat sau siguranță în funcționarea unui aparat).

Interpretarea rezultatelor se face prin:

1. Studierea tabelelor cu datele brute sau ordonate. Întrucît tabelele statistice sînt, de obicei, cu multe coloane și foarte multe cifre, cercetarea tabelelor se face doar pentru a depista unele greșeli de numerotare, de înregistrare etc., precum și pentru a stabili ordinul de mărime a numărului de elemente ce intră în calcule.

2. Studiul mediilor (aritmetice etc.) poate arăta dacă rezultatul este cu mult mai mare sau mai mic decît cel așteptat, deci dacă este logic sau nu. De asemenea, se mai pot vedea diferențele dintre diferitele medii (pe eșantioane, pe grupe de vîrstă etc.).

3. Studiul abaterilor patratice ne arată variabilitatea valorilor individuale față de medii: abatere mare=variabilitate mare. Tot aici, putem identifica valorile eronate ce trebuie îndepărtate. Repetarea cercetării după o pe-

Scanned with OKEN Scanner

rioadă de timp și obținerea unei alte abateri ne face să presupunem că populația s-a modificat. Cunoașterea compoziției populației (ex. dimensiunile corporale, a piciorului) sînt importante pentru planificarea activității din industria de confecții, industria încălțămintei etc., conform cu nevoile dimensionale ale populației. Pe baza unor eșantioane, se pot semnala la timp anumite schimbări, urmînd a se face o replanificare a activității de producție.

4. Compararea a două eșantioane se poate face cu testul χ^2, dar cînd avem elemente puține se întrebuințează testul t.

Scanned with OKEN Scanner

1

TABEL CU VALORILE CURBEI GAUSS-LAPLACE

σ	ordonata y	suprafața de sub $\frac{1}{2}$ curbă	σ	ordonata y	suprafața de sub $\frac{1}{2}$ curbă
0	0,3989	0,0000	1,1	0,2179	0,3643
0,05	0,3984	0,0199	1,2	0,1942	0,3849
0,1	0,3970	0,0398	1,3	0,1714	0,4032
0,15	0,3945	0,0596	1,4	0,1497	0,4192
0,2	0,3910	0,0793	1,5	0,1295	0,4332
0,25	0,3867	0,0987	1,6	0,1169	0,4452
0,3	0,3814	0,1179	1,7	0,0940	0,4554
0,35	0,3752	0,1368	1,8	0,0790	0,4641
0,4	0,3683	0,1554	1,9	0,0656	0,4713
0,45	0,3605	0,1736	2	0,0540	0,4772
0,5	0,3521	0,1415	2,25	0,0317	0,4878
0,55	0,3429	0,2088	2,5	0,0175	0,4938
0,6	0,3332	0,2257	2,75	0,0091	0,4970
0,65	0,3230	0,2422	3	0,0044	0,4986
0,7	0,3123	0,2580	3,25	0,0020	0,4994
0,75	0,3011	0,2734	3,5	0,00087	0,49977
0,8	0,2897	0,2881	3,75	0,00035	0,49991
0,85	0,2780	0,3023	4	0,00013	0,49996
0,9	0,2661	0,3159	4,5	0,0001	0,499997
1	0,2420	0,3413	5	0,00001	0,4999997

Scanned with OKEN Scanner

2

TABEL CU VALORILE t

p	0,9	0,7	0,5	0,3	0,2	0,1	0,05	0,01	0,001
f=1	0,158	0,510	1,000	1,953	3,078	6,314	12,70	63,65	636,61
2	0,142	0,445	0,816	1,386	1,886	2,920	4,303	9,92	31,59
3	0,137	0,424	0,765	1,250	1,638	2,353	3,182	5,84	12,94
4	0,134	0,414	0,741	1,190	1,533	2,132	2,776	4,60	8,61
5	0,132	0,408	0,727	1,156	1,476	2,051	2,571	4,03	6,87
6	0,131	0,404	0,718	1,134	1,410	1,943	2,447	3,70	5,96
7	0,130	0,402	0,711	1,119	1,405	1,895	2,385	3,49	5,40
8	0,130	0,399	0,706	1,108	1,397	1,860	2,306	3,35	5,04
9	0,129	0,398	0,703	1,100	1,383	1,833	2,282	3,25	4,78
10	0,129	0,397	0,700	1,093	1,372	1,812	2,228	3,16	4,58
12	0,128	0,395	0,695	1,083	1,356	1,782	2,782	3,05	4,31
14	0,128	0,393	0,692	1,076	1,345	1,761	2,145	2,97	4,14
16	0,128	0,392	0,690	1,071	1,337	1,746	2,120	3,92	4,01
18	0,127	0,392	0,688	1,067	1,330	1,734	2,101	2,87	3,92
20	0,127	0,391	0,687	1,064	1,325	1,725	2,086	2,84	3,85
22	0,127	0,390	0,686	1,061	1,321	1,717	2,074	2,81	3,79
25	0,127	0,390	0,684	1,058	1,316	1,708	2,060	2,78	3,72
30	0,127	0,389	0,683	1,055	1,310	1,697	2,040	2,75	3,64
40	0,126	0,388	0,681	1,050	1,303	1,684	2,021	2,70	3,55
60	0,126	0,387	0,679	1,046	1,296	1,671	2,000	2,66	3,46
120	0,126	0,384	0,677	1,051	1,289	1,658	1,98	2,61	3,37
∞	0,126	0,384	0,674	1,036	1,282	1,645	1,96	2,57	3,29

Scanned with OKEN Scanner

TABEL CU VALORILE χ^2

p	0,9	0,7	0,5	0,3	0,2	0,1	0,05	0,01	0,001
f=1	0,015	0,14	0,45	1,07	1,64	2,71	3,84	6,63	10,8
2	0,211	0,71	1,39	2,41	3,22	4,61	5,99	9,21	13,8
3	0,584	1,42	2,37	3,67	4,64	6,25	7,81	11,3	16,3
4	1,06	2,19	3,36	4,88	5,99	7,78	9,49	13,3	18,5
5	1,61	3,00	4,35	6,06	7,29	9,24	11,1	15,1	20,5
6	2,20	3,83	5,35	7,23	8,56	10,6	12,6	16,8	22,5
7	2,83	4,47	6,35	8,38	9,80	12,0	14,1	18,5	24,3
8	3,49	5,53	7,34	9,52	11,0	13,4	15,5	20,1	26,1
9	4,17	6,39	8,34	10,7	12,2	14,7	16,9	21,7	27,9
10	4,87	7,27	9,34	11,8	13,4	16,0	18,3	23,2	29,6
11	5,58	8,15	10,3	12,9	14,6	17,3	19,7	24,7	31,3
12	6,30	9,03	11,3	14,0	15,8	18,5	21,0	26,2	32,9
13	7,04	9,93	12,3	15,1	17,0	19,8	22,4	27,7	34,5
14	7,49	10,8	13,3	16,2	18,2	21,1	23,7	29,1	36,1
15	8,55	11,7	14,3	17,3	19,3	22,3	25,0	30,6	37,7
16	9,31	12,6	15,3	18,4	20,5	23,5	26,3	32,0	39,3
17	10,1	13,5	16,3	19,5	21,6	24,8	27,6	33,4	40,8
18	10,9	14,4	17,3	20,6	22,8	26	28,9	34,8	42,3
19	11,7	15,4	18,3	21,7	23,9	27,2	30,1	36,2	43,8
20	12,4	16,3	19,3	22,8	25	28,4	31,4	37,6	45,3

Scanned with OKEN Scanner

TABEL CU VALORILE χ^2 (CONTINUARE)

p	0,9	0,7	0,5	0,3	0,2	0,1	0,05	0,01	0,001
f=21	13,2	17,2	20,3	23,9	26,2	29,6	32,7	38,9	46,8
22	14	18,1	21,3	24,9	27,3	30,8	33,9	40,3	48,3
23	14,8	19	22,3	26	28,4	32	35,2	41,6	49,7
24	15,7	19,9	23,3	27,1	29,6	33,2	36,4	43	51,2
25	16,5	20,9	24,3	28,2	30,7	34,4	37,7	44,3	52,6
26	17,3	21,8	25,3	29,2	31,8	35,6	38,9	45,6	54,1
27	18,1	22,7	26,3	30,3	32,9	36,7	40,1	47	55,5
28	18,9	23,6	27,3	31,4	34	37,9	41,3	48,3	56,9
29	19,8	24,6	28,3	32,5	35,1	39,1	42,6	49,6	58,3
30	20,6	25,5	29,3	33,5	36,3	40,3	43,8	50,9	59,7
40	29,1	34,9	39,3	44,2	47,3	51,8	55,8	63,7	73,4
50	37,7	44,3	49,3	54,7	58,2	63,2	67,5	76,2	86,7
60	46,5	53,8	59,3	65,2	69	74,4	79,1	88,4	99,6
70	55,3	63,3	69,3	75,7	79,7	85,5	90,5	100,4	112,3
80	64,3	72,9	79,3	86,1	90,4	96,6	101,9	112,3	124,8
90	73,3	82,5	89,3	96,5	101,1	107,6	113,1	124,1	137,2
100	82,4	92,1	99,3	106,9	111	118,6	124,3	135,8	149,4

Scanned with OKEN Scanner

TABEL CU VALORILE t
PENTRU COMPARAREA MEDIILOR ARITMETICE

p	0,90	0,95	0,975	0,99	0,999
f=4	1,533	2,132	2,776	3,747	7,173
5	1,476	2,015	2,571	3,365	5,893
6	1,440	1,943	2,447	3,143	5,208
7	1,415	1,895	2,365	2,998	4,785
8	1,397	1,860	2,306	2,986	4,501
9	1,383	1,833	2,262	2,821	4,297
10	1,372	1,812	2,228	2,764	4,144
11	1,363	1,796	2,201	2,718	4,025
12	1,356	1,782	2,179	2,681	3,930
13	1,350	1,771	2,160	2,650	3,853
14	1,345	1,761	2,145	2,624	3,787
15	1,341	1,753	2,131	2,602	3,733
16	1,337	1,747	2,120	2,583	3,686
18	1,330	1,734	2,101	2,552	3,611
20	1,325	1,725	2,086	2,528	3,552
25	1,316	1,700	2,060	2,485	3,450
30	1,310	1,697	2,042	2,457	3,385
40	1,303	1,684	2,021	2,423	3,307
50	1,298	1,676	2,009	2,403	3,262
60	1,296	1,671	2,000	2,390	3,232
80	1,292	1,664	1,990	2,374	3,195
100	1,290	1,660	1,984	2,365	3,174
200	1,286	1,653	1,972	2,345	3,131
500	1,283	1,648	1,965	2,334	3,106
	1,282	1,645	1,960	2,326	3,090

Scanned with OKEN Scanner

TABEL CU VALORILE FISHER LA p=0,95

f_1 =	1	2	3	4	5	6	8	10	20	50	100	∞
f_2=1	161	200	216	225	230	234	239	242	248	252	253	254
2	18,5	19	19,2	19,2	19,3	19,3	19,4	19,4	19,4	19,5	19,5	19,5
3	10,1	9,55	9,28	9,17	9,01	8,91	8,85	8,79	8,66	8,58	8,55	8,53
4	7,71	6,94	6,59	6,39	6,26	6,16	6,04	5,96	5,80	5,70	5,66	5,63
5	6,61	5,79	5,41	5,19	5,05	4,95	4,82	4,74	4,56	4,44	4,41	4,37
6	5,99	5,14	4,71	4,53	4,59	4,28	4,05	4,06	3,87	3,75	3,71	3,67
7	5,59	4,74	4,35	4,12	3,97	3,87	3,73	3,63	3,44	3,32	3,28	3,23
8	5,32	4,46	4,07	3,84	3,69	3,58	3,44	3,34	3,15	3,03	2,98	2,93
10	4,96	4,10	3,71	3,48	3,33	3,22	3,07	2,97	2,77	2,64	2,59	2,54
14	4,60	3,74	3,34	3,11	2,96	2,95	2,70	2,60	2,39	2,24	2,19	2,13
20	4,35	3,49	3,10	2,87	2,71	2,60	2,45	2,35	2,12	1,19	1,90	1,84
24	4,26	3,40	3,01	2,78	2,62	2,36	2,36	2,26	2,02	1,86	1,80	1,73
30	4,17	3,32	2,92	2,69	2,53	2,42	2,27	2,16	1,93	1,76	1,69	1,62
40	4,08	3,23	2,84	2,61	2,45	2,34	2,18	2,07	1,84	1,66	1,59	1,51
50	4,03	3,18	2,79	2,56	2,40	2,29	2,13	2,02	1,78	1,52	1,50	1,44
100	3,94	3,03	2,70	2,46	2,31	2,19	2,13	1,92	1,68	1,48	1,39	1,28
∞	3,84	3,00	2,60	2,37	2,21	2,10	1,94	1,83	1,57	1,35	1,24	1,00

Scanned with OKEN Scanner

TABEL CU VALORILE FISHER LA p=0,99

$f_1=$	1	2	3	4	5	6	8	10	20	50	100	∞
$f_2=1$	405	500	540	563	576	586	598	606	621	630	635	637
2	98,5	99	99,2	99,2	99,3	99,3	99,4	99,4	99,5	99,5	99,5	99,5
3	24,1	30,8	29,5	28,7	28,8	27,9	27,5	27,1	26,7	26,4	26,2	26,1
4	21,2	18	16,7	16	15,5	15,2	14,8	14,5	14	13,7	13,6	13,5
5	16,3	13,3	12,1	11,4	11	10,7	10,3	10,1	9,55	9,24	9,13	9,02
6	13,7	10,9	9,87	9,15	8,75	8,74	8,10	7,89	7,40	7,09	6,99	6,88
8	11,3	8,65	7,59	7,01	6,63	6,37	6,03	5,83	5,36	5,06	4,96	4.86
10	10	7,56	6,55	5,90	5,64	5,39	5,06	4,85	4,41	4,12	4,01	3,91
14	8,86	6,56	5,56	5,05	4,70	4,46	4,14	3,94	3,51	3,21	3,11	3,00
20	8,10	5,85	4,94	4,43	4,10	3,87	3,56	3,37	2,94	2,63	2,53	2,42
24	7,82	5,65	4,72	4,22	3,90	3,67	3,36	3,17	2,74	2,44	2,33	2,21
30	7,56	5,59	4,51	4,02	3,70	3,30	3,17	2,98	2,55	2,24	2,13	2,01
40	7,31	5,18	4,31	3,83	3,51	3,29	2,99	2,80	2,37	2,05	1,94	1,81
50	7,17	5,06	4,20	3,72	3,41	3,19	2,89	2,70	2,26	1,94	1,82	1,68
100	6,90	4,82	3,98	3,51	3,21	2,99	2,69	2,51	2,06	1,73	1,59	1,43
∞	6,63	4,01	3,78	3,32	3,02	2,80	2,51	2,32	1,87	1,52	1,36	1,00

Scanned with OKEN Scanner

TABEL CU VALORILE TESTULUI GRUBBS

p =	0,95	0,98	0,99
f = 3	1,41	1,41	1,41
4	1,71	1,72	1,73
5	1,92	1,96	1,97
6	2,07	2,13	2,16
7	2,18	2,27	2,31
8	2,27	2,37	2,43
9	2,35	2,46	2,53
10	2,41	2,54	2,62
11	2,47	2,61	2,69
12	2,52	2,66	2,75
13	2,56	2,71	2,81
14	2,60	2,76	2,86
15	2,64	2,80	2,91
16	2,67	2,84	2,95
17	2,70	2,87	2,98
18	2,73	2,90	3,02
19	2,75	2,93	3,05
20	2,78	2,96	3,08

TABEL CU VALORILE TESTULUI Q (DEAN ŞI DIXON)

p =	0,90	0,95	0,99
f=3	0,89	0,94	0,99
4	0,68	0,77	0,89
5	0,56	0,64	0,76
6	0,48	0,56	0,70
7	0,43	0,51	0,64
8	0,40	0,48	0,58

Scanned with OKEN Scanner

TABEL CU VALORILE TESTULUI ROMANOVSKI

p =	0,95	0,98	0,99
f=3	4,93	8,04	11,46
4	3,56	5,08	6,53
5	3,04	4,11	5,04
6	2,78	3,64	4,36
7	2,62	3,36	3,96
8	2,51	3,18	3,71
9	2,43	3,05	3,54
10	2,37	2,96	3,41
11	2,33	2,89	3,31
12	2,29	2,83	3,23
13	2,26	2,78	3,17
14	2,24	2,74	3,12
15	2,22	2,71	3,08
16	2,20	2,68	3,04
17	2,18	2,66	3,01
18	2,17	2,64	3,00
19	2,16	2,62	2,95
20	2,15	2,60	2,93

Scanned with OKEN Scanner

TABEL CU VALORILE TESTULUI λ (IRWIN)

p =	0,95	0,98	0,99
f=3	1,79	2,17	2,90
4	1,64	2,05	2,75
5	1,51	1,93	2,60
6	1,39	1,81	2,45
7	1,31	1,69	2,30
8	1,24	1,57	2,16
9	1,20	1,51	2,09
10	1,18	1,46	2,03
11	1,14	1,43	2,00
12	1,11	1,41	1,97
13	1,09	1,39	1,94
14	1,07	1,37	1,91
15	1,06	1,35	1,88
16	1,05	1,33	1,86
17	1,04	1,31	1,84
18	1,03	1,29	1,82
19	1,03	1,28	1,81
20	1,03	1,27	1,80

Scanned with OKEN Scanner

TABEL CU NUMERE ÎNTÎMPLĂTOARE

(extras din STAS 6085-72)

6744	5149	1898	9784	7522	5329	1052
2967	3599	4888	4068	6368	8339	3847
8331	2559	8748	2580	2408	8145	2132
4035	7295	4856	6319	8016	7757	4852
0924	9171	2976	5401	5341	2067	6295
7988	5791	1169	1022	7187	2492	5624
1918	3540	2766	8972	2117	7169	9517
9817	3733	9375	1888	3585	4605	0720
5751	4041	7624	1854	6061	7913	9457
7730	8330	5928	7333	4317	6007	4538
7062	4346	0613	2238	3118	6460	0794
3055	6747	9579	6366	8256	6710	7677
0989	6887	7919	3594	6618	1794	7281
1828	0888	4856	4944	6782	7267	2883
5551	2925	7295	1518	2568	4892	8716
8715	1523	1379	6273	7669	0977	8265
9961	9808	3411	8879	0832	4676	3358
5750	8580	5339	0592	5442	2901	3523
5355	8312	9502	7911	4974	8795	9848
6107	4905	4620	3578	3134	2450	6811
6915	6442	9239	3608	5639	3502	9278
9866	4909	9910	6253	1931	8183	5043
5954	8272	4466	6403	7659	1212	4156
4654	5138	5907	6421	8117	8841	2305
8279	9262	4424	0134	6415	3356	1779
5870	6836	4937	0825	5128	4209	4528
3101	5897	5610	9697	3275	6573	0601
8490	1991	7541	5563	9825	0969	2957
8799	9848	3813	6418	7850	4397	0407
9951	2713	4321	0755	9035	9084	0870

Scanned with OKEN Scanner

TABEL PENTRU STABILIREA VOLUMULUI EŞANTIOANELOR

p =	0,90	0,95	0,98	0,99	0,999
s=1,0	5	7	9	11	17
0,5	13	18	25	31	50
0,4	19	27	34	46	74
0,3	32	46	64	78	128
0,2	70	99	139	171	277
0,1	273	387	545	668	1 089
0,05	1 084	1 540	2 168	2 659	4 338

TABEL PENTRU STABILIREA VOLUMULUI EŞANTIOANELOR

n	p
5	0,960
6	0,970
7	0,976
8	0,980
9	0,983
10	0,985
12	0,988
14	0,990
16	0,991
18	0,992
20	0,993
25	0,994
30	0,995
50	0,996
150	0,997
∞	0,9973

Scanned with OKEN Scanner

Bibliografie

Alecsescu, M. *Atenţiune, vorbeşte luna.* Buc., Ed. Enciclop., 1970.

Adămuş, Z., Anghel, I. *Statistică teoretică.* Buc., Ed. Did. şi Ped., 1971.

Bobancu, V. *Caleidoscop matematic.* Buc., Ed. Albatros, 1979.

Baciu, I. *Fiziologia.* Buc., Ed. Did. şi Ped., 1970.

Bontilă, Gh. *Aptitudinile şi măsurarea lor.* Buc., 1971.

Baron, T., Korka, M., Pecican, E., Stănescu, M. *Statistică.* Buc., Ed. Did. şi Ped., 1981.

Biji, E., Porojan, D., Wagner, P., Pintilie, L. *Statistică.* ASE, 1982.

Ciorănescu, N. *Matematici pentru statisticieni.* Institutul de Statistică, 1943.

Ciucu, Gh., Craiu, V. *Introducere în teoria probabilităţilor şi statistică matematică.* Buc., Ed. Did. şi Ped., 1971.

Ceauşescu, D. *Utilizarea statisticii în chimia analitică.* Buc., Ed. Tehnică, 1980.

Craiu, V., Mihoc, G. *Matematica pentru economişti.* Buc., Ed. Tehnică, 1970.

Craiu, V. *Verificarea ipotezelor statistice.* Buc., Ed. Did. şi Ped., 1972.

Dumitrescu, Gh. *Metode şi tehnici matematice în pedagogie.* Buc., Ed. Did. şi Ped., 1969.

Giosan, N., Săulescu, N. *Principii de genetică.* Buc., Ed. Ceres, 1972.

Ionescu, H. *Valori medii.* În: „Revista de matematică-fizică", nr. 3, 1953.

Lineva, V., Osipova, L., Tamarina, N. *Metode de determinare a rezistenţei insectelor.* În: „Referate igienă", nr. 6, 1951.

Mureşan, P. *Metode matematice în clinică, laborator şi ocrotirea sănătăţii.* Buc., Ed. Medicală, 1976.

Scanned with OKEN Scanner

Breviar de statistică medico-sanitară. Min. Sănătății, 1967.

Moreaux, Th. *Influences astrales.* Paris, Ed. Doin, 1942.

Iosifescu, M., Moineagu, C., Trebici, V., Ursianu, E. *Mică enciclopedie de statistică,* Buc., Ed. Șt. și Enciclop., 1985.

Panaite, V., Munteanu, R. *Control statistic și fiabilitate.* Buc., Ed. Did. și Ped., 1982.

Pracontal S. A. *Une forme de cancer traité par les mathematiciens* in Science et vie, marte 1982.

Râmneanțu, P. *Elemente de biometrie medicală și statistică vitală.* Institutul de Statistică, 1939.

Rancu, N., Tövissi, L. *Statistica matematică cu aplicații în producție.* Buc., Ed. Academiei, 1974.

Rădulescu, V. *Revanșa minții.* Buc., Ed. Militară, 1974.

Săhleanu, V. *Metode matematice în cercetarea medico-biologică.* Buc., Ed. Med., 1957.

Sachter, O. *Șiruri recurente.* În: „Rev. matematică-fizică", nr. 10, 1952.

Shaw, B. *Aforisme, paradoxuri, cugetări.* Buc., Ed. Albatros, 1983.

Steinbach, M. *Prelucrarea statistică în medicină și biologie.* Buc., Ed. Academiei, 1961.

Trifan, N. *Curs de puericultură.* IMF Iași, 1975.

Tövissi, L., Vodă, V. *Statistică teoretică.* Buc., Ed. Șt. și Enciclop., 1982.

Tövissi, L., Baron, T., Cușa, G. *Controlul statistic al mărfurilor,* ASE, 1971.

Văleanu, I., Melinte, C. *Lucrări practice de igienă comunală,* IMF Iași, 1970.

Văleanu, I. *Curs de igienă școlară.* IMF Iași, 1970.

Văleanu, I. *Radioactivitatea locuinței.* Teză de doctorat, IMF Iași, 1970.

Văleanu, I., Meliniuc, V., Văleanu, M. *Considerații igienico-sanitare asupra cariei dentare.* USSM CFR, Constanța, 1978.

Văleanu, I., Hîncu, M. *Considerații asupra meteorosensibilității,* USSM CFR, Constanța, 1981.

Weiling, F. *Zum 100 Todestag vom Johann Gregor Mendel.* În: Forum Mikrobiologie, nr. 2, 1984.

Vodă, V. *Gîndirea statistică,* Buc., Ed. Albatros, 1977.

Scanned with OKEN Scanner

Voiculescu, C., Mihancea, N. *Virusuri și reactivitate.* Craiova, Ed. Scrisul Românesc, 1982.

Vasiliu, Fl. *Pe meridianul Yamoto. Buc.*, Ed. Sport-Turism, 1980.

Crișan, I. *Statistica analizelor chimice.* În: *Metode fizice de analiză.* Ministerul Petrolului, 1970.

* * * *Kleine Enziklopedie*, Leipzig, 1954.

Colecția *STAS.*

Colecția *Science et vie.*

Colecția Revista *Magazin.*

Colecția Revista *Știință și Tehnică.*

Colecția Revista *Tehnium.*

The Guinness Book of Records, Anglia, 1988.

Scanned with OKEN Scanner

Cuprins

Prefață 5

Capitolul 1: **Introducere. Istoric.** 7

Capitolul 2: **Noțiuni introductive în matematica statistică** . 11

2.1. Scrierea cifrelor 11
2.2. Șiruri (serii) de numere 12
2.3. Logaritmi 14
2.4. Mașini de calcul 17
2.5. Mașini electronice de calcul 18
2.6. Noțiuni de informatică 21

Capitolul 3: **Calcule statistice** (în colab. cu ing. Rovența Vasile) 25

3.1. Medii 25

3.1.1. Media aritmetică 25
3.1.2. Media geometrică 30
3.1.3. Media patratică 34
3.1.4. Media armonică 36
3.1.5. Media cronologică 38
3.1.6. Media glisantă 40
3.1.7. Media patratelor 41
3.1.8. Media progresivă 43
3.1.9. Media logaritmică 43
3.1.10. Media mediilor 44
3.1.11. Interpretarea rezultatelor și concluzii 45

3.2. Abaterea patratică 47
3.2.1. Interpretarea rezultatelor și aplicații practice . 53
3.3. Curba (clopotul) lui Gauss-Laplace 57
3.4. Variabilitatea mediilor 65

3.4.1. Eroarea Standard ES sau $\sigma_{\bar{x}}$ 66
3.4.2. Eroarea Probabilă EP sau x_1 70
3.4.3. Logica utilizării valorii ES și EP 70

3.5. Testul t 71
3.6. Testul x^2 74

Capitolul 4: **Comparații** 84

4.1. Compararea mediilor aritmetice 84
4.2. Compararea abaterilor patratice 85

Scanned with OKEN Scanner

4.3. Coeficientul de variație 85
4.4. Indici 86

4.4.1. Indicele simplu 86
4.4.2. Indicele procentual 87
4.4.3. Indicele cronologic 87
4.4.4. Nivelul mediu procentual 88
4.4.5. Entropia informațională 88

Capitolul 5: **Eliminarea valorilor eronate** 90

5.1. Testul celor „3 sigme" 90
5.2. Testul Grubbs 91
5.3. Testul Q (Dean și Dixon) 92
5.4. Testul t 92
5.5. Testul Romanovski 92
5.6. Testul λ (Irwin) 93
5.7. Interpretarea rezultatelor 94

Capitolul 6: **Eșantionul** 95

6.1. Alegerea eșantionului 95
6.2. Selectarea eșantionului 104
6.3. Stabilirea volumului eșantionului 106

Capitolul 7: **Probleme rezolvate** 108

7.1. Obținerea de informații dintr-o singură valoare . . 108
7.2. Obținerea de informații din două valori 108
7.3. Șir progresiv geometric 109
7.4. Medii 110
7.5. Abaterea patratică 112
7.6. Testul t 113
7.7. Testul χ^2 113
7.8. Stabilirea gradului de libertate la textul χ^2 117
7.9. Problemă soluționată prin mai multe metode . . . 120
7.10. Alegere preferențială sau selectare opțională 123

Încheiere 124

Tabele 126

Bibliografie 139

Scanned with OKEN Scanner

REDACTOR: **MAGDALENA VASILESCU**
TEHNOREDACTOR: **FLORICA PREDA**
BUN DE TIPAR: 2 DECEMBRIE 1989
COLI DE TIPAR: 9
LUCRARE APĂRUTĂ ÎN REGIA AUTORULUI.

Tiparul executat sub cda. 307/1989, la
Întreprinderea Poligrafică „Crișana", Oradea,
str. L. Sălăjan nr. 105.
România

Scanned with OKEN Scanner